

NASA Contractor Report 177639

Effects of Checklist Interface on Non-verbal Crew Communications

Leon D. Segal, Ph.D.

Western Aerospace Laboratories, Inc.
1611 Mays Avenue
Monte Serrano, CA 95030

Prepared for
Ames Research Center
CONTRACT NCC2-486
May 1994



National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035-1000

Acknowledgments

Much of my thinking about problems involving the nature of behavior and communication has been inspired by the ideas and work of Gregory Bateson; the clear and systematic manner with which he makes sense of mind and nature served as a sharpening tool for my mind, often dulled by the long and rigorous academic process. I have been similarly inspired by the work of John Coltrane; his relentless explorations of patterns, along with his artistic integrity, are to me a continuous source for creative energy.

Throughout the process of producing this experiment and resultant report, Chris Wickens helped focus my thoughts and actions. I have not ceased to be impressed by his ability to discern information from noise and his gift for packaging ideas in words and pictures. He is truly an artist; as apprentice, I was privileged to practice my skills under his guidance

Barbara Kanki first read my "non-verbal" ideas when I sent that original report to NASA Ames Research Center in 1987. From that time on, she encouraged me to continue my line of investigation and to move to California in order to do my research as part of her "Crew Factors" group at NASA ARC. The crew of researchers she has assembled around her were instrumental in the completion of my work; I feel indebted to the "coding buddies," especially, to Kimberly Jobe and Mike Wich, who had the patience and stamina to stick with me until I found a way to coherently and effectively present and apply my ideas.

This research was conducted under grant NCA 2-616 from NASA-Ames University Consortium and supported by the Federal Aviation Administration and Western Aerospace Laboratories, Inc. Dr. Barbara Kanki served as technical monitor and chief collaborator for NASA ARC and Prof. Chris Wickens as academic advisor at the University of Illinois at Urbana-Champaign.

Table of contents

1. Introduction	1
1.1. General	1
1.1.1. Background research: ASRS reports	2
1.2. Team-machine interface: fundamental characteristics	6
1.2.1. Crew communication	7
1.2.2. Perception and redundancy	10
1.2.3. Remote vs. same workspace	12
2. Proposed perspective	17
2.1. Consequential communication	17
2.2. Actions as context for speech interpretation	19
2.3. Team Engagement State Space (TESS): a descriptive tool	20
2.4. Design as choreography	22
2.5. Summary of introduction and literature review	27
3. Methodology	29
3.1. Background: the Palmer and Degani study	29
3.1.1. Equipment	29
3.1.2. Experimental objectives	30
3.1.3. Independent variable: checklist interface design	30
3.1.4. Subjects	33
3.1.5. Simulated scenario	33
3.2. Results of preliminary analyses	36
3.3. Hypotheses	40
3.4. Proposed data interpretation and coding	40
4. Results	45
4.1. Non-verbal activity	45
4.2. Verbal utterances	53
4.3. Relationship between verbal and non-verbal	55
4.4. Performance	57
5. Discussion	62
5.1. Non-verbal activity	62

5.2. Speech Acts	67
5.3. Performance	69
5.4. Team Engagement State Space (TESS) - a qualitative interpretation	71
5.5. Potential applications and future research	74
6. Summary	79
References	80

Abstract

The investigation described hereunder looked at the effects of the spatial layout and functionality of cockpit displays and controls on crew communication. Specifically, the study focused on the intra-cockpit crew interaction - and subsequent task performance - of airline pilots flying different configurations of a new electronic checklist, designed and tested in a high-fidelity simulator at NASA Ames Research Center. The first part of this proposal establishes the theoretical background for the assumptions underlying the research, suggesting that in the context of the interaction between a multi-operator crew and a machine, the design and configuration of the interface will affect interactions between individual operators and the machine, and subsequently, the interaction between operators. In view of the latest trends in cockpit interface design and flight-deck technology - in particular, the centralization of displays and controls - the introduction identifies certain problems associated with these modern designs, and suggests specific design issues to which the expected results could be applied. A detailed research program and methodology is outlined, and the results are described and discussed. Overall, differences in cockpit design were shown to impact the activity within the cockpit, including interactions between pilots and aircraft and the cooperative interactions between pilots.

1. Introduction

1.1. General

Advances in technology have greatly impacted the appearance of the modern aircraft cockpit. Where once one would expect to see rows upon rows of dials and switches, the modern cockpit resembles a rack of CRT screens and data-entry keyboards. Technological advances have also changed the aviation system as a whole: more sophisticated and accurate air-traffic control facilities guide highly automated aircraft through highly congested airways, often in marginal weather conditions, on their way to airports that are operating at close to maximum capacity. With the accelerating speed of technological development, particularly with the miniaturization of high-speed computers that so greatly impacts the weight-conscious aviation industry, the trend toward added cockpit automation is expected to continue.

The rapid pace of introduction of computer-based devices into the cockpit has outstripped the ability of designers, pilots, and operators to formulate an overall strategy for their use and implementation (Wiener, 1988). Within this technological frenzy, at the center of the expanding network of sophisticated technologies, the human operators - the pilots - find themselves continuously trying to keep up with the fast rate of change. Not surprisingly, a large percentage of aircraft accidents are attributed to human error: for air carriers, about two thirds of all accidents are attributable to the cockpit crew; for general aviation, the rates are even more disproportionate, with almost 9 out of 10 accidents attributable to human causes (Nagel, 1988). While these figures may accurately reflect the apparent cause of an accident, it is important to remember that, rather than being the main instigators of accidents, operators tend to be the inheritors of system defects created by poor design, incorrect installation, faulty maintenance and bad management decisions (Reason, 1990). This research was directed at the first of these inheritances with which the pilots have to deal, namely, the design of cockpit interface.

How does the design of a cockpit affect the performance of each pilot operating within it? How do differences in individual performance affect the synergetic performance of the pilots as a crew? How does cockpit design impact on crew communication? This study looked at the constraints imposed on crew communication by the design of the system which they operate, focusing on the airline cockpit as one context in which design and communication are viewed as linked. Building on the analysis of 'real world' accounts, and

on theoretical constructs concerning human communication and human-machine interaction, a research program is described. The primary objective of the research was to identify the relationship between cockpit configuration and crew communication style - in particular, the non-verbal aspect of crew communication - with the applied purpose of defining useful guidelines for future cockpit-design processes.

As will be described in greater detail below, this project was based on the assumption that in the interaction between a team of operators and a machine, much of the information exchanged between operators is non-verbal. Further, it was assumed that the particular design of the interface between team and machine may affect the non-verbal aspect of information exchange, and thus directly impact the way in which the team-members communicate and coordinate their activities. Finally, this research program was oriented toward application; the only kind of research that qualifies as human factors research is research that is undertaken with the aim of contributing to the solution of some design problem (Chapanis, 1991). Accordingly, the Discussion (section 5.) looks at the potential applications of conclusions drawn from the results obtained in this study.

In order to establish the need for this research program, the author searched through existing records for indications that crew communication does indeed rely on the flow of activity-related, non-verbal, information for the engagement and coordination of individual operators. In particular, the search focused on incidents in which operators themselves described the situation as one in which they had used information other than verbal in performing their task within the crew context. This report begins with a discussion of several examples in aviation which provide the backdrop for the following theoretical discussion and subsequent description of the research program.

1.1.1. Background research: ASRS reports

In the preparatory phase of this study, accounts were compiled from NASA's Aviation Safety Reporting System (ASRS, 1991), a database that holds reports submitted voluntarily by aircrew following incidents which they consider irregular. The reports are subjective accounts of what transpired in the cockpit during the incident, as narrated by operators who were actively involved in the flight process. At the time of this search, the ASRS database contained 35,151 full-form records received since January 1, 1986. There were also 67,987 abbreviated-form records in the database, but since these do not contain textual information, they were excluded from the search. In all, 424 relevant reports were identified. The real

power of ASRS lies in the report narratives, as the reporting system is limited from a statistical perspective: since all ASRS reports are voluntarily submitted, they can not be considered a measured random sample of the full population of like events. Moreover, since not all pilots are equally aware of the ASRS, the data reflect reporting biases. As suggested by the ASRS office, only one thing can be known for sure from ASRS statistics - they represent the lower measure of the true number of such events which are occurring.

While reading through the brief reports, it is helpful to reconstruct the situation, and imagine the sequence of events unfolding. Readers should notice the particular impact of non-verbal information on crew communication and coordination. To aid with this perspective, each example is followed by the author's remarks, which were prepared based on the author's extensive aviation experience, as well as discussions of these examples with other experienced pilots. The examples have been numbered, as they will be used to illustrate points made in the discussion further below. The accounts have been slightly edited to allow for better reading.

Example 1: "The warning horn indicating cabin altitude exceeding 10,000' came on while cruising at 35,000'. Before I could silence the horn and analyze the problem... the First Officer commenced and executed an emergency descent drill: throttles to idle, speed-brakes out, and transponder frequency to 7700 (emergency signal) - exactly by the book. The aircraft was well out of 35,000' before I could comprehend that an emergency descent was in progress. No discussion, just down we go! What happened to crew coordination? I have no answer..."(ASRS #137152)

Remarks: Note the shared context provided by the auditory information from the warning horn, a context within which the Captain (CA) was able to make sense of what the First Officer (FO) was doing (even though he did not necessarily agree with the actions). Notice the level of detail at which he was able to perceive the FO's actions; he did not need the FO to tell him *what* he was doing - that was obvious from seeing the actions; nor did he need to be told *why* it was being done - that was obvious from hearing the horn. The only thing missing, from the perspective of the reporting operator, was a discussion of whether that particular procedure, rather than any other, was appropriate.

Example 2: "As we added power for takeoff... the Second Officer (SO) pointed out a light aircraft immediately ahead of us at about 1000'... I almost decided to abort the

takeoff, but a second later the light aircraft began a right turn, so I continued the takeoff. Because of the disturbance and worry about the light aircraft, I did not notice or keep track of the rapidly accelerating airspeed, and missed the FO's callout of takeoff speeds. Meanwhile, because of the increasing speed, the airplane began to pitch up and the FO, sensing that we were tail-heavy, began to trim the stabilizer nose-down... I was unaware that the FO had changed the stabilizer setting, and therefore misunderstood when the takeoff warning horn sounded (it sounds automatically when the trim is beyond a certain setting). I began to reduce power to abort the takeoff, but both FO and SO correctly advised me that it was too late to abort, and that the trim had been changed. Takeoff was continued with no further problems." (ASRS #96238)

Remarks: Notice the different sources of information to which operators were directing their attention: the CA was engaged with looking at the light aircraft, and missed the FO's calls for takeoff speed; the FO was engaged with the change in pitch, and missed the connection between that and the increasing speed; the CA was unaware that the FO had changed the stabilizer setting, and thus misinterpreted the warning horn. Notice that since both FO and SO knew that the setting was changed, they understood that the horn was warning about the change in stabilizer setting (trim out of the green band), and, realizing that the CA was unaware of their high speed and that the trim had been changed, correctly advised him to continue the takeoff.

Example 3: "On instrument approach to landing. The aircraft was well above the glide-slope. At approximately 700' above ground level, the First Officer advised of the need for a missed approach procedure. Go around was called, power was applied by Captain, gear and flaps were raised by First Officer; no verbal commands were given, and no acknowledgments of actions were made. The aircraft immediately pitched up to excessive nose high attitude, well in excess of 35 degrees. The airspeed bled off to approximately 80 knots. At this point, the Captain was unable to exert enough forward control pressure and yelled for First Officer to help. Both pilots had full forward control pressure with the stick shaker sounding (stall warning). Neither pilots were able to take their hands off the controls. The Captain twice yelled for flaps up; after second call the Flight Engineer raised flaps in response to Captain's command. The nose of the aircraft began gradually lowering to less excessive pitch. After critical stage, the climb and approach checklist were then read again, and an uneventful approach and landing was made..."(ASRS #79837)

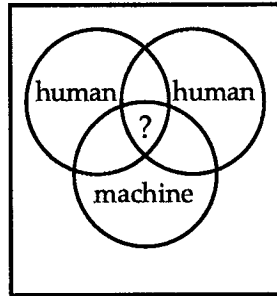
Remarks: Once the go around had been announced, i.e. clear intention verbally stated, the actual control actions were performed silently; there was so much redundant information that indicated that engine power has been applied - e.g., throttles forward, instrument readings, increase in engine sound, acceleration felt by the "seat of the pants" - there was no need for the CA to say "power applied." There was, however, *not* enough information about the true flap position - although FO reached for, and manipulated, the flap control, he apparently moved it to the wrong position. Throughout the entire event, the aircraft's extreme and unusual attitude provided all three crew members with a context within which actions were perceived and interpreted. Note the reference to the stick shaker, which provided both tactile and auditory information that warned of an imminent stall. Now, in contrast to this very dynamic situation, notice how informative the *lack* of action may be:

Example 4: "...While all of this was happening, I noticed the Captain was unresponsive. I know that he was diet controlled diabetic and told him to eat. He did, and became responsive in a few minutes..." (ASRS #84100)

Remarks: Within this particular control context, observing a *lack* of activity prompted the FO to draw certain conclusions, and take corrective action to remedy this potentially disastrous situation. Thus, while actions provided the context for interpretation of events in the first three examples, *inaction* was the key piece of information in this account.

The accounts illustrate the point that, in the context of a team workstation such as an aircraft cockpit - where experts cooperate in performing a highly-demanding task - it is almost impossible for an individual operator to do anything that does *not* constitute inter-crew communication. Actions performed by one operator, or expected actions that are not performed, carry certain expectancies and meanings for other operators (see Klein, 1989, for a discussion of the close connection between expertise and expectancies). Thus, actions have consequences not only for the machine, but also for other operators who are present when the actions are performed. The study of team coordination, and the design of team-machine interface, must therefore include in their scope more than simply the auditory, speech-based, interaction between team members.

1.2. Team-machine interface: fundamental characteristics



In sports, being a good team player involves more than simply being proficient at the technical skills of the game and standing on the same court with other players. In basketball, for example, some players are considered good individual players, while others are considered good team players. Since basketball is fundamentally a team sport, and at the same time, one where players often find themselves facing one-on-one situations, the combination of these two attributes - good individual *and* team skills - was what made great players such as Magic Johnson, Larry Bird, or Michael Jordan.

When teams are assembled - whether in sports, business, industry, or any other area of cooperative (*co-operative*) human activity - the context created by the aggregation of operators has unique attributes. Above all, it defines two tasks, two categories of interactions, pertaining to two activities that the operators need perform; two different - yet inevitably integrated - classes of interaction. One class of interaction emerges from the *interface between the team and the task environment* in which they operate; in this category, we look at the team's overall goals, and how team actions affect the machine and the environment. The other class of interaction emerges from the *interface between individual team members*; here, we focus on how the flow of information within the team affects the coordination and cooperation between individual members. This *group process* is shaped by the interplay of aspects of its composition and structure, and the demands of the task environment (McGrath, 1964). Thus, in the particular context of team-machine interaction, aggregating operators creates a context that includes two categories of interactions: that which is usually called "control," and that which is called "communication." The first category includes the exchange of information between operators and machine through displays and controls; the second category includes the exchange of information between operators, utilizing all modalities of perception and action.

The arguments developed in the following discussion outline the critical role played by designers of team-machine systems in determining the particular relationship between the two tasks of control and communication. Since the group process includes the interactions among members as the group performs its task within its environmental setting (McGrath, 1964), it seems reasonable to assume that significant changes to that environmental setting could result in significant changes in the group process; technological characteristics have influences on individuals, as well as on organizational structures (Hulin & Roznowski, 1985). While acknowledging that team cooperation and performance is affected by operator training, experience and skill level, investigation of personal aptitude, learning and expertise goes beyond the scope of this research program.

1.2.1. Crew communication

Crew cooperation (co-operation) within a specific task environment depends directly on the quality of communication between the individual crew members. The dictionary defines "communication" as: "The act of imparting, conferring or delivering, from one to another" (Webster's, 1983). What crew members continuously exchange in the cockpit is information, and it is on the crucial flow of information that this discussion focuses.

The questions have been asked before: What is the nature of the mechanism that allows a group to work together in a task performance situation? How is responsibility delegated, how often do group members communicate with each other, what is the style of their interchange? (Foushee & Helmreich, 1988). Numerous experimental and conceptual papers have made significant contributions towards answering these questions (for several important studies of crew interaction in the aviation domain, see: Foushee & Manos, 1981; Foushee, 1982, 1984; Kanki et al., 1989a; Kanki et al., 1989b; Linde & Shively, 1988; Costley et al., 1989; Shaffer et al., 1988; Foushee & Helmreich, 1988; Strauss and Cooper, 1989). Current understanding of crew coordination stands on the broad base of empirical and theoretical data accrued by these and other studies. These studies have revealed some important, systematic effects relating the frequency and type of communication to the efficiency - or error-proneness - of cockpit crew performance.

Foushee & Manos' (1981) analysis of cockpit voice recordings uncovered several interesting relationships between crew communication and performance. While there was a tendency for crews who communicated less not to perform as well, the type or quality of communication played an even more pivotal role in performance: there was a negative

correlation between crew member observations about flight status and certain type of system-control errors. At the same time, there was a negative correlation between system-control errors and acknowledgments of information provided by other crew members; increased acknowledgments were also related to fewer errors overall. Frequency of commands was associated with a lower incidence of flying errors. Finally, crews who tended to make more errors exhibited higher rates of response uncertainty, frustration or anger, embarrassment, and lower rates of agreement.

The two studies by Kanki et al. (1989a; 1989b) used a high-fidelity full-mission simulation, flown by active line pilots, to study the relationship between communication variations and aircrew performance. The realistic flight scenario included weather conditions and an equipment failure which necessitated a substantial amount of crew coordination. Performance was evaluated on-line by expert observers present in the simulator cab during the sessions, as well as by later analysis of videotapes. Subsequently, crews were categorized in terms of high, mid, or low-error descriptions. Transcriptions of all verbal exchanges were coded into a system which focused on 2-step speech sequences: initiating speech (commands, questions, observations and dysfluencies), and response speech (reply, acknowledgments and zero response). Although the identification of specific patterns was not clear, overall, the pattern which distinguished the 4 low-error (high performance) crews from the rest was the adoption of a standard form of communicating. The researchers concluded that coordination and performance were enhanced by the regularities in communications which afforded predictability of crewmember behavior and confirmation of expectations.

Linde & Shively (1988) conducted a field study of the operations of police helicopters, focusing on the relationship between workload and crew communication. The data were collected during two weeks of audio and video recordings of in-flight police operations; subjective ratings of workload were taken in flight, and during debriefing. In addition, the heart rates of both crew members were recorded. They found that the frequency of intra-cockpit communication depended on the segment of the flight (mission vs. transit), and on communication type (mission relevant vs. non-mission relevant). Thus, when the workload was high, virtually no time was spent on non-mission relevant communication; conversely, when the workload was low, the crew "filled" the auditory space by discussing non-mission relevant issues. Further analysis of the transcripts was performed, in an attempt to identify the usefulness of linguistic fluency as a measure of crew fatigue. Their results suggest that a fatigued crew was 2.5 times more likely to commit a linguistic dysfluency - such as a stutter

or hesitation - than a rested crew. They concluded that the voice channel is not "free" of variations in workload, and that implementations of voice-activated input/output systems should take this into consideration.

Strauss & Cooper (1989) attempted to identify the impact of automation and crew structure on crew communication and performance. In their simulation study, 50 two-person crews flew a ninety minute mission, in which they encountered seven failures of ground-based navigational aids and cockpit automated equipment. Crews were composed to be either heterogeneous or homogeneous with respect to flight experience, age, leadership style, and attitudes toward crew interaction. These two levels of crew composition were crossed with two levels of aircraft automation - no automation, or two axis autopilot. Analysis of the transcribed intra-cockpit communication suggest that under conditions of automated control, more information exchange took place. Crews that exchanged a higher ratio of task relevant to task irrelevant communication performed better in these circumstances. Although the communication patterns were often related to performance in predictable ways, the investigators were unable to conclude any causal relationship between the two, and suggest that further analysis is necessary.

Most studies of air-crew communication have focused uniquely on the verbal interaction between the operators, analyzing recorded speech transcripts in an effort to capture the information flow within the crew. The above dictionary definition of "communication," however, does not limit itself exclusively to speech, i.e., the act of communication is not inherently confined to verbal interaction. Further, as suggested by aviation lore (e.g. Hawkins, 1987), and with the support of such documented accounts as the ASRS records presented above (1.1.1.), pilots do capitalize on more than verbal/auditory information for crew coordination. By default, this added category of information gets labeled as "non-verbal." Interestingly, while the 1970's marked a transition in the way social sciences regard non-verbal communication, moving toward a legitimate and identifiable area of scholarship (Burgoon, 1980), this perspective has yet to be embraced by the engineering-oriented human factors community. Thus, while the amount of information pertaining to non-verbal communications has become overwhelming in the "softer" social sciences (Burgoon et al., 1989), information pertaining to non-verbal interactions within a particular task environment - specifically, in the context of mutli-operator interactions with a continuous control tasks such as flying - is almost non-existent. It seems reasonable to assume it unlikely that - given the availability of several sources of information - pilots would actually rely on verbal information alone; whenever possible, humans, like all living organisms, rely

on redundancy for perception. The discussion in the next section makes a slight detour from the aviation paradigm in order to further establish the role of redundancy - and multiple sources of information - in perception and action.

1.2.2. Perception and redundancy

Perception and redundancy are intimately linked; perceptual systems - i.e. living organisms - learn to capitalize on the redundancy of information available in their environment. Further, organisms learn to capitalize on redundancy itself as a source of higher-level information ("higher" in relation to individual events), i.e. on the patterning or predictability of particular events within a larger aggregate of events (Birdwhistell, 1970; Bateson, 1972). For a good illustration of this point, we can look to nature, which always provides wonderful examples of the ecology of behavior.

A small European bird, the indigo bunting, migrates at night, relying primarily on the pattern of stars in the sky to orient itself. Although it is sensitive to the earth's magnetic field, when the sky is visible, the indigo bunting will use the visual information even if it conflicts with the sensory information provided by the magnetic field. In a fascinating study, Emlen (1975) demonstrated that the bunting orients itself relative to that particular point in the night sky around which all stars rotate; thus, the bird inevitably directs itself in relation to the north, or polar, star. By placing the birds in a planetarium and rotating the night sky around stars other than the north star, Emlen produced a tendency in the birds to orient in the direction of the artificial polar star; even if the point of singularity, the fixed point, was occluded by clouds, the birds were able to accurately orient themselves toward it by referring to the angular motion of the stars that were visible. In this scenario, redundancy exists within the same modality - i.e. all relevant information is detected by the visual system. The redundancy of information available in the visually rotating field of stars, and the organism's ability to capitalize on that redundancy, are key to understanding the indigo bunting's behavior, i.e. the dynamic organization of its perception and action.

The ability of pigeons to navigate accurately during day or night is well documented, and has been exploited by man for many centuries. Experimental studies suggest that pigeons are able to use the sun's position in the sky as a directional guide (Keeton, 1974); on overcast days, or at night, they use their sensitivity to the earth's magnetic field to do the same (Wiltschko et al., 1981). Interestingly, experimental disruptions of the bird's detection of the magnetic field (using a miniature electric coil fitted on its head, along with a battery

strapped to its back - quite a hideous sight) have no effect whatsoever when the sun is visible; they do, however, cause disorientation when the sun is occluded by clouds. Thus, it seems that the information provided by the sun's position preempts the information provided by the magnetic field (Walcott, 1972). Further complexity was introduced by Kiepenheuer (1985), who demonstrated that pigeons can also use olfaction as a means for orientation. These findings demonstrate not only the pigeon's ability to sense different forms of directional information, but also its ability to differentiate between the different information fields, and select the particular form that is most appropriate for any given context.

Like the indigo bunting and the pigeon, the human perceptual system too thrives on redundancy. For example, in the case of locomotion and navigation, information about the direction of motion is available redundantly in the entire optic flow field, in accelerations perceived by the vestibular system, in doppler-type effects produced by sounds bouncing off stationary objects, in kinesthetic feedback from limbs and muscles, and in shifts perceived by the olfactory system. While certain perceptual modalities are relied upon more than others - e.g., visual information usually plays a larger role in human navigation than does olfaction - there are situations in which the prioritization of information sources may change, or at least, situations in which the relationship between the information provided by two different modalities provides information at a higher level of analysis. For example, certain kinds of electrical fires may start with a strong smell of burning plastic, without any visual cues such as smoke or flames; the nose detects information that, along with the *lack* of visual information, suggests a particular type of event. Another good illustration of the well established role of redundant processing in the perception of depth, described by Wickens et al. (1989); in this review, motion, stereopsis and occlusion were determined to be particularly salient cues for depth perception. As in the case of motion, depth perception relies on the interaction between multiple cues - albeit in the same modality - the weight of which is determined by the particular context and task.

In communicating, humans rely on a broad spectrum of information provided by both words and actions, as well as by the environment within which the interaction takes place. Humans - just like the other living organisms described above - learn to capitalize on all forms of information that the situation presents (Rochlin, 1987; Lave, 1988; Hutchins, 1989; Klein, 1989). True, in certain cases, language and related symbol systems (e.g., American Sign Language, Morse Code) may be used to convey information that can only be presented symbolically, such as the description of a plan, the diagnosis of a situation, or discussing

yesterday's dinner. At most times, however, particularly when people are actively involved in a physical task and interact within the same physical environment, what people say is not independent of what they do. If we study communication by looking at verbal interactions alone, we set ourselves up, a priori, to learn about only a subset of the communication domain. In order to capture the complexities of behavior, we have to work with a model that allows us to see the complexities (Olsen & Rasmussen, 1989). Thus, it must be assumed that there is more to crew communication than meets the ear; in order to capture the complexity of this behavioral domain, crew communication should be investigated from the broadest perspective possible.

1.2.3. Remote vs. same workspace

Because of the critical relevance of workspace design to task structure, a distinction must be made between situations in which operators share the task of control but do so from different stations, and those situations in which operators control the system while sharing one, multi-operator control station. In the first situation, operators in individual, segregated - and often remote - stations, coordinate their activities *through* the system (the "machine"), using auditory and visual channels provided by the system to transmit and receive information; radios, telephones, screens displaying verbal and video information, as well as feedback pertaining to changes in the system, provide individual operators with the information essential for crew coordination. Individual operators interact directly - i.e., physically and visually - with the technological system only, and indirectly - i.e., via channels provided by the mediating technology - with the task environment, which includes other operators.

McGrath & Hollingshead (1993), in a discussion of Group Support Systems (GSS), make a similar distinction between "distal" and "contiguous" groups. As they accurately point out, while contiguous - i.e., face-to-face - communication may rely on all modalities, distal communication entails reliance on only a reduced set of modalities. Unfortunately, they include discussion of face-to-face communication only as a baseline for comparison, and do not expand on the potential impact of GSS that are used in contiguous, same workspace, situations. Although they acknowledge that technological enhancements do not necessarily require group members to be spatially separated from each other, their analysis focuses on the impact of technology on distal groups. The literature contains quite a number of related articles (e.g., Helander, 1985; McGuire et al., 1987; Olson & Olson, 1992; McLeon, 1992; Heath & Luff, 1992), most of which look at the extreme scenario of communication between

people operating from different locations. The contribution of these studies to the current discussion is in their description of group cooperation that is often completely devoid of direct interaction. From this perspective, one may imagine a group communication continuum, anchored at one end with direct face-to-face communication, and, at the other end, with pure technology-mediated communication, such as electronic mail. Thus, differences in cockpit automation may be creating operational environments that lie at different points along this continuum.

While cockpit automation does indeed introduce novel information into the group process, the fundamental quality of crew communication in the cockpit remains one of "content rich," (McGrath & Hollingshead, 1993), face-to-face communication. Thus, the information added to the cockpit environment by technology and automation is precisely that - added information, as opposed to GSS in the classical sense of the word, where technology aims to replace and compensate for lost information. A recently published article may serve to illustrate the essential components of GSS studies. Kiesler & Sproull (1992) looked at how people behave in decision-making meetings conducted through one particular electronic communications technology - electronic mail. They compared face-to-face meetings with the computer-mediated condition, in which each group member participated from a different office in the same building. Groups were presented with a decision task; raw data included individual and group choices, questionnaire response, and tape recordings and electronic transcripts of group discussions.

The results were powerful: it takes much longer for an electronically-mediated group to make a decision - as much as 10 times as long - as it does for the face-to-face group; in electronic discussions, participation is distributed more equally, regardless of status or gender; when groups make decisions using electronic communication, they have more difficulty reaching consensus than when they meet face-to-face. It seems reasonable to assume that these three findings are related: e.g. in the electronic communication group, greater participation leads to difficulty in reaching consensus, taking a longer time for decision-making. Interestingly, problems which were designed to measure the riskiness and quality of decisions yielded differences between the two conditions: groups that met face-to-face were risk averse for gain choices and risk seeking for loss choices, while groups that met using the computer to communicate were somewhat risk seeking in all circumstances. The investigators speculate that the basis of the risk-taking difference was the reduced social information - e.g. non-verbals, status, gender - which may have made group members less sensitive to implicit social pressure and accountability.

The researchers attribute all the differences between groups to one or another aspect of the attenuation - to at least some degree - of the social context cues available in face-to-face conversations. However, these social context cues are available in the cockpit; research has shown that personal attitude and group orientations had significant impact on group process variables and crew coordination (e.g. Helmreich, 1982). Thus, cockpit technology can be assumed to play a different role in shaping crew communication. In the context of the research described hereunder, it would be interesting to see whether any of these findings demonstrated by Kiesler and Sproull emerge in cockpit crew communication. Does cockpit automation affect the time it takes the crew to make a decision? Does it enable crew members to state their opinion more clearly, thus creating more balance and participation in the decision process? What role does automation play in the riskiness of decision-making? While some of these questions go beyond the particular objectives of this proposal, it is hoped that the results will provide an initial basis for the discussion and comparison of GSS and cockpit automation.

In contrast with the distal GSS paradigm described above, when operators share the same workstation they are exposed to both face-to-face *and* technological information; when sharing a workstation, every team-member has access not only to environmental and system information, but also to the information provided *directly* by the presence and "co-operation" of team-members. In this situation, the physical layout of the system creates a context where each operator's activities can be directly perceived by other operators, and thus can directly inform and affect team coordination. This shared situational context (Nickerson, 1981) - or physical co-presence (Gibbs & Muller, 1990) - has been noted to affect verbal communication in social settings, and to structure team coordination tactics in goal-oriented task performance activities. One example of GSS research that gets closer to this scenario is provided by a study of interpersonal video-mediated communication by Heath and Luff (1992), in which mutual visual access was shown to provide individuals with the ability to discern the ongoing organization and demands of a colleague's activities and thereby to coordinate their interaction with the practical task at hand. The researchers argue that mutual visual access provided individuals with the ability to point at and refer to objects within the shared local milieu.

In a study that looked at both distal and contiguous communication scenarios, Chapanis et al. (1972) investigated the communication between two operators who's task it was to assemble a commercial product according to written instructions provided by the manufacturer. The teams consisted of two operators - one who read the instructions,

another who did the actual assembly work - cooperated under different conditions of information exchange, ranging from written notes only, through verbal exchange from two different rooms, to co-presence in the same room (what they called: "the communication-rich mode"). The task did not require any particular skill - the operators were high-school students, who met only for the purpose of the experiment. As the contact between the operators became more direct, the frequency and amount of communication between them increased. Further, when the two operators were allowed to work in the same room, the modality of communication changed: "In the communication-rich mode, subjects could, and did, use non-verbal forms of communication. They gestured, nodded, grimaced, and used other expressive movements of the body." Unfortunately, the investigators did not have the tools to analyze this aspect of team coordination. As they describe it: "In some preliminary trials, video-tape recordings were made to see if such non-verbal forms of communication could be quantified. After consulting the literature and after reviewing the tapes, no practical way of quantifying what was seen was apparent and the attempt was abandoned."

Rochlin et al. (1987), who studied patterns of team cooperation on board aircraft carriers, note that the physical co-presence of operators - coupled with the practice of verbal announcement and verification of actions - create an organization with a remarkable degree of personal and organizational flexibility. They note that almost everyone involved in bringing the aircraft on board is part of a constant loop of conversation and verification; seasoned personnel do not listen so much as monitor for deviations; in the Combat Decision Center, a number of people are "just watching," keeping track of each other's jobs, or monitoring the situation from other locations. Operation manuals rarely discuss integration of tasks into the whole; team coordination tactics are structured by the situation, and by the individual, micro-level rules and details of specific tasks.

In a study of teamwork in performing the task of maritime navigation, Hutchins (1989) looked at the effects of technology on operator communication and crew coordination. He found that when the navigation task was performed by the team, the coordination among the actions of the members of the team was not achieved by following a master procedure; instead, it emerged from the interaction among members of the team, and their interaction with the available technology. Coordination among the activities of the team members arose because some of the conditions for each team member's actions were produced by the activities of the other members of the team; this coordination tactic he calls "coordination by mutual constraint." Thus, the actions performed by each operator, and the verbal exchange

between operators - much of which was often "overheard" or monitored by others - these provided the structure around which crew coordination was established.

In summary, in the context of crew-type tasks, crew members base their understanding of the "state of the world" on all the information that is available to them, i.e. the aggregate of both verbal (speech) and non-verbal (actions) behavior, as well as the information provided by the technological environment. Verbal communication uses only one of the many kinds of signals that people can exchange; for a balanced view of the communication process we should always keep in mind the great variety of other signals that can reinforce or contradict the verbal message (Miller, 1973). In the communication process, much information is conveyed by context and nonlinguistic means (Nickerson, 1981). Verbal and non-verbal communication systems are redundant and complementary (Haslett, 1987). Communication within a shared physical space - e.g. crew communication within a cockpit - is, by necessity, a synergy of both verbal and non-verbal information.

2. Proposed perspective

2.1. Consequential communication

The particular class of communication which emerges from cooperative human activity in situations of co-presence is defined in this dissertation as "consequential communication." In this definition, consequential communication is a by-product of one operator's goal-driven activity; it is detected by the observer, rather than intentionally broadcast by the sender. While the classic definition of "communication" suggests some intentionality on part of the sender - "...imparting, conferring or delivering, from one to another" (Webster, 1983) - consequential communication emerges from the purposeful interaction between human and environment, rather than purposeful interaction between humans. Here, the actor is more of an emitter than a sender, and the observer, more of a perceiver than a receiver (John Grinder, personal communication). The perceiver is responsible for the categorization of the relevant information as "communication" - i.e., whether the event is to be categorized as an act of communication depends upon the perceptual skills and directed attention of the perceiver. Since when a word is spoken, all those who happen to be in perceptual range of the event will have some sort of "participation status" relative to it (Goffman, 1981) - i.e., the perceiver of a word becomes, almost by definition, the listener - we may similarly assume that when an action is performed, those who observe it will have some participation status relative to it.

Consider the information available to a passenger in the front right-hand seat of a car, watching the driver navigate through morning traffic. Imagine that you are sitting there, glancing at the driver to your left - what information can you detect? One source of information is the environment, i.e. the surrounding vehicles and objects as seen through the windshield and the different parameters in the car's displays. The other source of information is the driver's actions - his physical interaction with the car's displays and controls, his reactions to the information he perceives in the environment: If he slams on the brakes, he most probably noticed an obstacle or impending danger; if he operates his turn signal, he will soon be changing lanes; very frequent glances up at the rearview mirror suggest a tailgating car (perhaps a police car?); when he looks down to tune the radio he is attending to its auditory feedback, and stands in danger of missing important information on the road; frequent glances at street signs suggest he might be lost, or looking for a specific address; and so on.

The above example illustrates several ways in which non-verbal, consequential communication can be used as an effective mode of cooperative coordination. This may serve as a method of monitoring individual performance and providing feedback essential for team structure (Larson & LaFasto, 1989). Table 1 describes how consequential communication fits into the global domain of "communication."

One principle holds for all cooperative task performance: once the workspace has been laid out and the spatial context for physical actions defined, any perceivable, observed behavior may serve for communication (Segal, 1989). As long as the observing operator is sufficiently familiar with the task and its workspace, he can - and will - assign meaning to all actions taken within that workspace. Actors in specific situations expect technically competent role performance from each other (Barber, 1983); that expectation may be met, or not, on the basis of what they see each other do. Even the lack of behavior - i.e., no observable action - carries meaning, which, in ASRS Example 4 (1.1.1.), communicates: "I have not eaten enough." In this sense, operators sharing a workspace cannot "not communicate" - the common physical context for actions inherently ascribes meaning to any and all observable behavior. Non-verbal information plays a critical role in creating the context for the crew's verbal communications - sometimes replacing these communications altogether - and must therefore be addressed in concurrence with verbal information.

	intentional	unintended
verbal	direct dialogue	while CA talks on the radio, FO listens
non-verbal	gestures, signs	task performance provides consequential communication

Table 1: categories of verbal and non-verbal communication

2.2. Actions as context for speech interpretation

Without context, words and actions have no meaning at all (Bateson, 1979); the information conveyed is not an intrinsic property of the individual message (Ashby, 1956), but rather, emerges from the interaction between event and context. Utterance and situation are bound up inextricably with each other, and the context of situation is indispensable for the understanding of the words (Malinowski, 1923). Sometimes what a person is saying unconsciously by his actions may directly contradict what he is saying consciously with his words (Miller, 1973). Actions provide a context that plays a critical role in the interpretation of verbal communication, as suggested by various researchers (Nickerson, 1981; Kreckel, 1981; Haslett, 1987; Gibbs & Muller, 1990). Speech communication takes place within the physical environment of the workspace, and is not independent of the physical behavior that continuously takes place within the same environment (Segal, 1989). Thus, a copilot who reacts appropriately to a command to do 'X' need not actually say "Roger, I am doing 'X'." A copilot who says: "Got you," then proceeds to do something other than what the captain intended, is in fact communicating: "I did not understand your command." A pilot who sits immobile at the controls, visually fixated with a glazed look on one instrument only and mumbling: "Don't worry, everything is under control," may in fact be communicating that things have gotten out of control.

The importance of contextual information provided by non-verbal, physical behavior is most evident in recent studies pertaining to the implementation of "intelligent" machines in team positions that were heretofore occupied by humans. In particular, the aviation-oriented study by Small et al. (1988) highlights the crucial role for a context shared by the operators as a basis for understanding verbal communication. In an attempt to define the prerequisites which would enable the use of a "pilot's associate" - i.e. an automated, AI-driven replacement for the copilot, designed on the basis of auditory interface - the investigation focused on the essential characteristics of verbal interactions in the cockpit. While the interface design assumed pilot inputs based on the technological capability of fast and accurate speech recognition, the limitations inherent in such human-machine interactions were very apparent. As described by the investigators in their concluding remarks: "The single most important message from the Cockpit Natural Language study is that the knowledge context (goals, intentions and situational awareness) of the pilot is crucial to understanding the semantics of the (verbal) commands."

The relationship between context and the interpretation of speech is a pivotal point in the study of team communication and coordination. The physical form of the workstation, the task structure, and the environment within which the team-machine system interacts - these define a hierarchical set of nested contexts, and thus constrain the range of information that may be perceived and understood by the operators. It is incumbent upon the designers of complex systems to remain aware of the role they play in defining the context within which the different issues described above - activities of "control" and "communication," individual operator's perception of redundant information, and the emergence of consequential communication - take place.

2.3. Team Engagement State Space (TESS): a descriptive tool

When studying a situation such as described above, i.e., where humans act and perceive in physical co-presence, it is often useful to envision a two-dimensional space, one which contains all the states which describe the mutually constraining relationship between the two. The Team Engagement State Space, or TESS (Figure 2), defines a two dimensional space which can be used to describe the team's engagement in the two different tasks, their direction of attention to the two different sources of information offered by the team context. It presents "communication" as one dimension of information, one class of interactions; it is this class of information that, as discussed above, defines the difference between single-operator and multi-operator contexts. The other dimension, "control," refers to the class of information that is specifically related to the interface between operators and system, i.e., the vertical axis describes the operators' engagement in the task of interacting with, and controlling, the system. Note that on both axes, the scale of 0-1 does not represent any specific, quantifiable measurement of information or behavior; its purpose is to illustrate, in an abstract way, the relative extent to which operators are engaged in the particular dimension described.

Imagine an aircraft parked at an airport, one hour before takeoff; while the cabin crew go about their duties, the pilots spend several minutes discussing the flight plan. In this situation, their involvement with the system is minimal, though their engagement with each other is high (point 'a'). Now think of that same crew performing the first portion of their pre-flight checklist: each operator is in charge of checking a different group of subsystems and paperwork; each is strongly engaged with the system, with very little exchange of information - i.e. neither intentional nor consequential communication - going on between them (point 'b'). Finally, picture the situation during the takeoff roll itself: the captain is

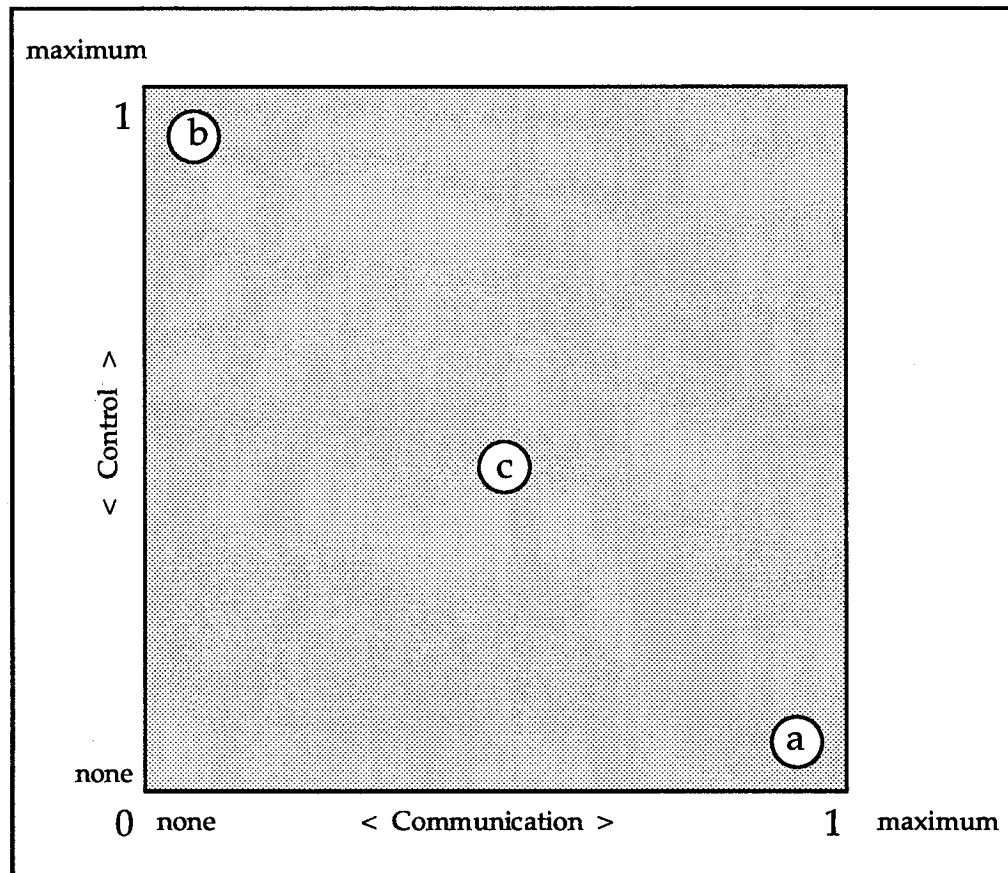


Figure 2: Team Engagement State Space (TESS)

controlling the aircraft, while the first officer announces the speeds and sets the throttles to takeoff thrust; as soon as the aircraft is airborne, the captain calls out: "Positive climb, gear up," upon which the first officer reaches for the control that raises the gear; the captain glances to confirm these actions. In this situation, both operators are engaged in control, as well as in both intentional and consequential communication (point 'c').

From an investigative point of view, if a particular team is observed as operating within a certain region of TESS, three primary, applied questions can be asked: What characteristics, or qualities, of *the team* makes it operate in this region? What elements of *the task* constrain the team to operate in this region? What properties of *the system* and *the environment* constrain the team to operate in this region? To this end, it would be interesting to note if different crews operating a similar system fall in the same area of the chart, or how changes in flight conditions might change a particular crew's location. Similar questions could be asked about individual operators, focusing on the relationship between different crew

members. From this perspective, one may look at the relative location of crew members in the TESS plot. Would control and communication change at the same rate for both operators? Could one operator's position change while the other's does not?

The primary objective in using TESS is to assist in clarifying some of the unique properties of the team-machine paradigm, namely, the confounded relationship between control and communication. Further, as a secondary objective, the TESS may provide a useful tool for visualizing and exploring the resultant data set, in a manner suggested by Sanderson (1991), after Tukey (1977).

2.4. Design as choreography

The transfer of information from the human to the machine will in general be in the form of manual actions on the system (Rasmussen, 1986); it is, after all, called human machine *interaction*, not human-machine *inter-speech* or *inter-thought*. While one needs to acknowledge the rapid development of speech-recognition technology and voice-activated controls, it seems that the bulk of control activity, particularly continuous control tasks, will remain physical. This being the case, the constraining relationship between the design of an environment and the behavior of living systems within that environment must be considered. Further, the applicability of the current study is of primary concern. As suggested by Chapanis (1991), whenever possible, human factors research should attempt to provide design recommendations; it is incumbent upon the researcher to make the connection between the theoretical basis, the findings, and application of results to material design problems.

The two different interfaces depicted in Figure 3 will serve to illustrate the immediate impact of display/control layout on operator behavior. Notice that, while both address the same interface problem - four controls for input, four displays for feedback - the difference in their design affords different styles of human-machine interaction. Imagine you were watching an operator interact with these interfaces: how would you perceive the effects of design on operator behavior?

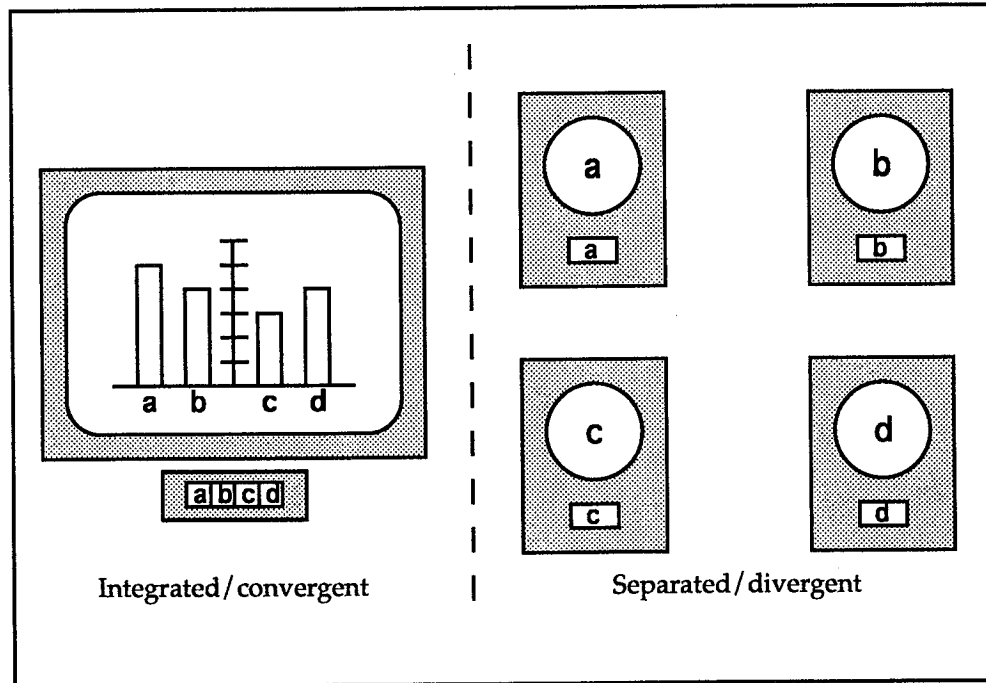


Figure 3: Two approaches to display/control layout

As proposed in the discussion above, the physical form of the work-station defines the constraints that *shape* operators' physical behavior; the spatial layout of the control environment constrains operators to perform a particular set of movements. Location of displays constrains them to direct their gaze and focus on particular points in space; location of controls constrains them to reach for those locations; the particular type of control constrains them to perform particular actions, e.g., push, twist, flip, pull. Using the sample layouts described in Fig. 3, one can imagine observing an operator monitor and control variables 'a' and 'd': notice the different movements of eyes, hand, and head, dictated by the two different layouts. If one were to supervise their performance, what layout would be preferred? Would that preference change if one were to operate the system oneself?

While physical form defines spatial constraints, the machine's operating procedures define and constrain the *temporal* organization of operators' behaviors; procedures that prescribe sequences of inputs impose on the operator constraints that dictate a particular pattern of actions over time. For example, a procedure might determine that control 'a' must never be manipulated before 'b'; control 'd' must always precede 'c', which is usually followed by 'b'. Such procedures, coupled with the system's response time, define a pattern of *actions over time* that is particular to every design, as well as to every context of operation. From this

perspective, the designer who defines the machine's logic of operation, the sequence of "if - then" statements that govern the system, is at the same time building a temporal sequence to which operators will conform in their interaction with the system. Again, Fig. 3 may serve to help imagine the difference between the patterns of actions emerging from interactions with the two different designs. From the perspective of an observing crewmember, note the differential impact of system layout on consequential communication.

Using this simple example, it is easy to see how designers are continuously confronted with tradeoffs throughout the design process. Since movements of the head that are needed to read instruments in a control room may take part in the identification of the source of the information (Rasmussen, 1986), integrated displays, while useful at alleviating the workload imposed on single operators (Wickens, 1992), may greatly reduce the availability of consequential communication, thus impairing the ability to monitor operator behavior. Integrated controls may similarly reduce one operator's abilities to observe another's activities. These considerations come into play when deciding on cockpit configurations for future aircraft, for example, in the current debate over the configuration of the cockpit for the High Speed Commercial Transport, where the narrow cross-section of the fuselage has caused a resurgence of the question: "Should the crew sit side-by-side, or in tandem?" (McDonnell Douglas Technical Report, 1992).

Currently, crew systems and their integration into the cockpit are being affected dramatically by new technologies, particularly increased on-board computer capabilities (Sexton, 1988). Elements such as CRT displays with multi-function capabilities and high graphic resolution enable integration and spatial centralization of displayed information; keyboards and control-mounted switches - e.g. Hands On Throttle And Stick (HOTAS) technology - provide spatial focal points for control inputs. The traditional control panel has been either partially or completely replaced by visual display units, and the traditional buttons, levers and knobs have been replaced by keyboards; this can create new problems (Ivergard, 1989).

Unfortunately, most contemporary design trends are driven primarily by the designer's fascination with state-of-the-art technology. When everything becomes possible, when all limitations are gone, design (and art) can easily become a never-ending search for novelty, until newness-for-the-sake-of-newness becomes the only measure (Papanek, 1985). The people responsible for the design of an environment are not always aware of its strong effect on the interactions that take place within that environment (Burgoon et al., 1989). Perrow

(1983) draws a distinction between "design logic" and "operating logic"; according to his perspective, these represent two different approaches to systems' design. He describes one contradiction between design logic and operating logic that is particularly relevant to the current discussion. In this example, he points out that while good design, by design logic, is compact, good operating logic stresses easy access to controls and system-state information. Thus, while good design favors single purpose information sources and controls, e.g. multi-function displays (MFD), good operation requires many entry points into the system for confirming information from different sources. While interface design is constrained by other considerations - e.g., space, weight, performance - once again we see the two conflicting performance criteria that designers attempt to balance - compact vs. accessible, single vs. redundant. This phenomenon - the conflict between design logic and operating logic - can be best described with the following "futuristic" example:

Imagine an aircraft that can be controlled by thought alone - a "utopian" dream entertained by many engineers. Assuming that all potential problems of measurement and reliability associated with such a system are solved, consider the difficulties such a design would impose on the cooperative work of a crew. The copilot would find it impossible to interpret any actions taken by the pilot, since no physical actions would be present for observation in the first place. In this type of control configuration, the only source of information pertaining to pilot or copilot actions would be the responses of the aircraft and its systems to the preceding control inputs. Thus, this design creates a critical phase lag in crew action information effectively ruling out any possibility for intervention in the case that one of the crew members makes a control or procedural error. Compare this to what happens in current designs where, when an error occurs (in the cockpit of an airliner), it is usually caught by one of the members of the crew and corrected immediately (Stone & Babcock, 1988). In our futuristic example, the crew environment constrains and inhibits the use of non-verbal cues to such an extent that it directly affects the form and quality of crew communication and cooperation.

One possible improvement could be the use of Multi Function Displays, where seeing which particular page an operator has selected may provide more specific contextual information to an observer. The level of detail in this case, however, would depend on the logical architecture of the menu-driven MFD: few pages of highly integrated displays would provide less consequential communication than many pages of fewer dimensions, though they may impose greater workload on the individual operator. The particular design of an

MFD becomes more critical as touch-sensitive screens serve not only as sources of information, but also as locations for control inputs. Since this research deals specifically with a touch-sensitive display and input interface device, it is hoped that the results shed some light on the role of interface design in shaping crew coordination.

ASRS reports provide some powerful examples that illustrate the impact of particular designs on aircrew activities:

Example 5: "I observed the FO switching frequency but I could not see the frequency selected in the window because the FO's hand covered it due to its location next to the autopilot turn knob... I reached down and raised the FO's hand off the turn knob and observed (that he had entered the wrong frequency)..." (#59918).

Example 6: "At this time I looked at my new FO's radio panel and saw he was going to transmit on the wrong channel, so I reached over and punched some buttons to put him on the right one... (#56215).

Example 7 "(During emergency procedure for shutting down engine) captain put his hand on #2 stop and feather and called "#2 stop and feather." I looked down to verify his hand position and called "pull." Captain continued with the engine fail checklist, and as he continued with the non memory items, I verified his hand on the proper controls and made calls per the checklist. While watching captains hand, we (passed the airport at which we were suppose to land).. then landed at wrong airport... (#105775).

Throughout the design process, the designer must be aware of these two principles: the *system's physical form shapes operator behavior, while its operating procedures organize that behavior temporally* (Segal, 1990). Designers must see themselves as choreographers; ideas that emerge from the designer's drafting table will define a set of actions unfolding over time, an operating "dance" that will be performed by operators whenever they interact with that system. It is incumbent upon the designer that the dance performed allow for the smooth flow of information between team members, and result in task performance that is effective, productive, and safe.

2.5. Summary of introduction and literature review

As a result of the rapid advancement of cockpit technology, along with the general increase in automation in the aviation domain, aircrews are operating in environments that are very different to those which they faced merely fifteen years ago. Meanwhile, the introduction of automation and computer-based technology has not relieved pilots of their ultimate responsibilities; the majority of air-traffic accidents are still attributed to breakdowns in pilot and crew performance. It was hypothesized that one possible impact of cockpit automation is in the area of crew communication; this hypothesis was supported with discussion of research and theory from several different fields. Group theory (e.g., McGrath, 1964; Hulin & Roznowski, 1985; Gibbs & Muller, 1990; McGrath & Hollingshead, 1993) provided the theoretical support for the importance of the role of technology and communication in the performance of the task. Experimental and observational studies from different domains, including aviation (e.g., Foushee & Manos, 1981; Kanki et al., 1982; Rochlin et al. 1987; Hutchins, 1989; Strauss & Cooper, 1989), provided empirical support for the argument that crew communication and task performance are linked. In the emerging domain of Group Support Systems (GSS), the data suggest that technology indeed alters the dynamics between group members (Kiesler & Sproull, 1992; Heath & Luff, 1992). With the resurgence of tandem cockpits, such as the suggested configuration of the future High Speed Civil Transport, the knowledge accrued through GSS studies may become particularly relevant for the understanding of cockpit design and crew communication.

It was suggested that non-verbal information plays an important role in group communication. This was supported by theoretical discussions, observational and experimental studies (e.g., Malinowski, 1923; Birdwhistell, 1970; Chapanis et al., 1972; Miller, 1973; Nickerson, 1981; Burgoon et al., 1989; Heath & Luff, 1992). The proposed connection between automation and non-verbal communication, and the proposal of the concept of "consequential communication," were partially supported by numerous real-world examples taken from NASA's Aviation Safety Reporting System (ASRS). With the support provided by the above mentioned theoretical and empirical studies, in light of the reports supplied by aircrew via the ASRS, and with the growing trend of cockpit automation, it seemed essential to start collecting empirical data concerning the immediate impact of cockpit automation on pilot task performance, and its subsequent impact on crew communication and coordination.

Finally, the notion of consequential communication and the TESS (Figure 2) were presented as conceptual tools for defining the problem, and as aids for discussing possible applications. These served as the basis for the critical distinction between two types of operator behavior - control and communication - and demonstrated the close relationship between the two in the context of team-machine interaction.

The current investigation was designed to tie these areas together, through an attempt to answer two questions: 1. "Do pilots use non-verbal information for task coordination?" and 2. "How does cockpit interface impact the non-verbal communication between crewmembers?" In the simulator study described in detail in the sections below, twelve two-pilot crews flew the same flight scenario using three different types of checklist interface. All flights were recorded on video tape, and crew performance was evaluated by an expert observer during the flight. Subsequently, non-verbal and verbal cockpit activities were coded and transcribed from the video tapes, and additional performance ratings were given by other expert observers. The video transcripts and performance measures served as the basis for data analysis. Details of the research methodology, results and discussion of the results are presented below.

3. Methodology

3.1. Background: the Palmer and Degani study

High fidelity flight simulation is extremely expensive, both in direct cost, as well as in the amount of time it demands from skilled professionals. Further, while the recruitment of expert pilots greatly increases the validity of the findings, it is very difficult to obtain access to this unique subject pool. For these reasons, many simulator studies combine the work of several different investigators; as long as experimental manipulations do not actually conflict with one another, the same raw data may serve different investigators to study different aspects of crew performance and behavior. This study used video data recorded in the winter of 1990 as the basis for investigation of the non-verbal paradigm.

The original simulation, designed by Ev Palmer and Asaf Degani - FLT Branch, Aerospace Human Factors Division (ASHFRD) at NASA-Ames Research Center - manipulated particular interface issues that are highly relevant in the context of the current discussion. Their interest, however, focused on the errors performed by the crews; the video tapes have thus far been used only for error coding, not for the study of crew interaction or non-verbal communication cues. Their initial findings were presented in a brief paper (Palmer and Degani, 1991); an in-depth technical report has not yet been released (Palmer and Degani, in preparation). Another study (Mosier, 1992) made use of the data to look at crew problem resolution, but again, did not use the video data to investigate the particular variables involved in crew communication.

3.1.1. Equipment

The experimental simulation study was carried out in the Advanced Concepts Flight Simulator (ACFS), at the Man Vehicle System Research Facility (MVSRF) at NASA's Ames Research Center. The ACFS was designed to simulate a two-person flight crew, twin turbofan, advanced airliner with the capacity to carry approximately 200 passengers. The simulator is run by a VAX 8030 main computer, which generates the flight dynamics and concurrent visual display. Although the simulator is mounted on hydraulic pylons that afford full motion, this capability was not used in this study.

The ACFS served as the test-bed for an electronic checklist interface design and evaluation. Cockpit avionics included five color CRT displays (10.5" x 13.5"), which displayed primary

and secondary flight instrumentation, systems' status information, and the two different types of electronic checklist for the appropriate experimental conditions (see Figure 3, section 3.1.3. below). Pilots could manipulate the information and systems displayed through a touch panel overlay, which was mounted over each display. Thus, in the so called "electronic" checklist conditions, performance of the checklist entailed repeated reaching for, and touching of, the CRT displays.

3.1.2. Experimental objectives

The objective of the Palmer and Degani study was to investigate the effectiveness of electronic checklists for commercial air transport. Three different checklist systems were designed and evaluated in the ACFS. The checklists were designed with the objective of reducing errors, and the investigators' primary research questions focused on the effects of checklist design on crew errors. Since these three checklists served as the independent variable in the current research program, it is important to discuss their design and functionality in greater detail.

3.1.3. Independent variable: checklist interface design

The 12 crews that participated in this experiment flew identical scenarios, using three types of checklists (4 crews per experimental condition, or checklist type). Of the three different checklists, two were designed as new, electronic checklists; the third checklist used was the standard paper checklist, which served in the original study as the control condition. The "Normal Procedures" portion of this checklist was printed using standard airline format on a single 8.5" x 11" paper card, and the irregular procedures bound in a booklet modeled after Boeing's Quick Reference Handbook format. Typically, the paper checklist is hand-held by the crew-member responsible for reading out the relevant items. While this freed these pilots from the need to interact with the checklist display described below, they still had to interact with aircraft systems, such as fuel, electric and hydraulic systems. Those interactions were performed with a graphical interface provided on the top half of a touch-sensitive CRT display, mounted in front of each pilot; the bottom half of that CRT displayed the electronic checklist.

The two electronic checklists were the Manual-Sensed Checklist (Manual), and the Automatic-Sensed Checklist (Auto). The interaction with both was similar, in that the pilots used the touch-sensitive CRT screens in the process of performing individual items in order

to complete the required checklist. The checklists could be presented on whichever CRT display faced the pilot-not-flying (PNF), that is, the pilot who was in charge of performing the checklist procedures. The checklist appeared in the bottom half of the screen, while aircraft system graphics and controls appeared in the top half. The particular selection of screen location was performed by the pilots themselves, according to their own distribution of flight duties. Note that, for any particular leg in the flight, either pilot could be Pilot Flying (PF) or Pilot Not Flying (PNF); the division of roles was left up to the crew itself, as is the case in routine flight operations. Figure 4 presents a sketch of the cockpit layout, with a sample checklist page. Note the location of the checklist at the bottom half of the CRT display.

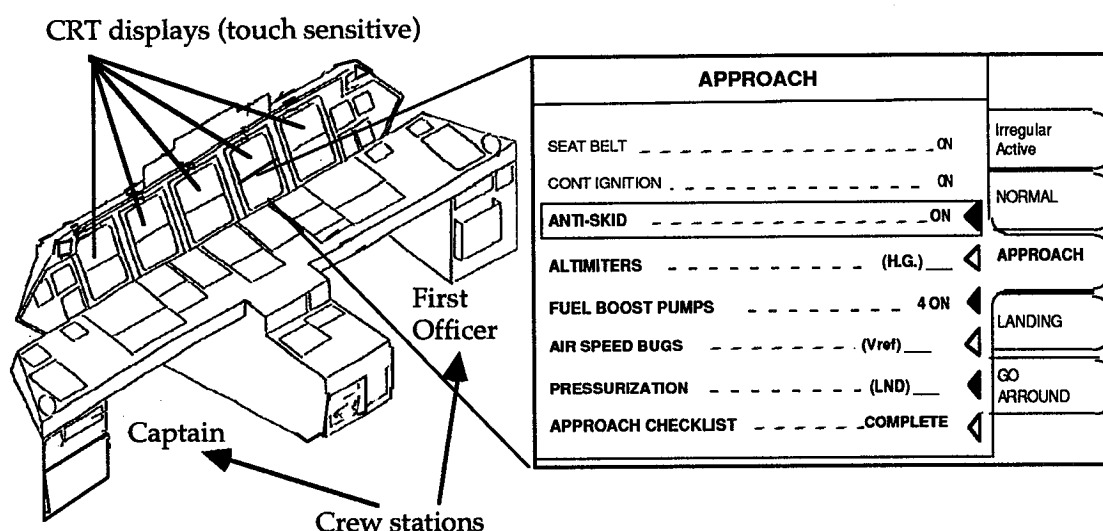


Figure 4: ACFS cockpit layout, and sample checklist page

Both electronic checklists relied on color coding for displaying item status: items-not-yet-performed were white, items-completed were green, and items-incomplete/skipped were amber. While both electronic checklists were similar in appearance and graphic design, they differed in the functional aspects of their interface; specifically, they differed in the level of acknowledgments required from the pilots. The Manual required the pilots to acknowledge - by touching the screen at the appropriate location - each item on the list of items to perform. Once an item was acknowledged by the pilot, the checklist system would send an interrogating signal to verify the state of that particular item, e.g., if "Flaps 25°" was touched by the pilot, the checklist system would verify if the flaps were indeed set at 25°. If the item were indeed "sensed" as completed, the checklist display would reflect that fact by scrolling

down to the next item. If the item were sensed as incomplete or skipped, it would turn amber. Once the last item on the checklist was touched, all uncompleted or skipped items would be displayed, upon which the pilots had to choose whether they wanted to complete them, or, if the situation warranted and authorization was given from airline operations, override them.

The Auto would require the pilot to interact only with those items which were not completed; as soon as the checklist was called up, all the items were automatically scanned, and those sensed as complete would be displayed in green. The items sensed as not completed, and the items that could not be sensed by the system, were displayed in white - e.g. "Cabin notification" is a checklist item which the system cannot check. Additionally, an indication of whether an item could be sensed was provided by triangles on each line: a dark triangle identified a sensed item, while a triangle outline identified an unsensed item (Fig. 4). While pilots operating the Manual needed to touch the screen for every single item included in the checklist, pilots in the Auto condition were required to check only those items which were sensed by the system as incomplete, or that could not be sensed. Specifically for the purpose of the current research, note that the Manual design required pilots to reach for, and manipulate, the checklist CRT more often than did the Auto design. Approximately 70% of the checklist items were sensed in the Auto, thus reducing the number of interactions pilots in that condition had to make with the display.

During the simulated flights, the following data were recorded:

1. Three infra-red video cameras were mounted above and behind the pilots, and recorded events in the cockpit. One wide-angle camera captured the entire cockpit, including the pilots' backs; it is this view which provided most of the non-verbal crew interaction information intended for use in the current proposal. The two other cameras were each focused on each pilot's primary flight displays; these cameras provided most of the information pertaining to the interaction between each individual pilot and the aircraft's systems.
2. Several system variables (alpha-numeric) were recorded, and displayed on a fourth videos screen in the experimenter operations station: time, a/c speed and heading, distance from way points, checklist type (experimental condition), and checklist status.
3. Two audio channels were used to record intra-cockpit communication, communication of the crew with Air Traffic Control (ATC) and ground crew, and ATC communications with other aircraft on the particular radio frequencies selected by the crew.

4. During the experimental sessions, an observer - a highly experienced pilot who had participated in previous simulation studies at NASA-Ames - sat in a jumpseat in the cockpit, observed all crews, and rated the pilots' individual performance at the end of each leg. At the end of the session, the observer rated the crews' overall performance. Performance ratings were filled on forms listing various qualitative performance categories, e.g., crew management style, checklist performance, workload distribution. For each category, the observer would rate performance on a scale of 1 to 5, representing "poor" to "excellent" performance, respectively.

3.1.4. Subjects

Twenty four pilots, all from the same major US carrier, were recruited by posters placed in the carrier's flight operations room. The recruitment process, which was arranged so that twelve two-person crews could be put together, was coordinated with the airline's management, as well as with representatives of the Air Line Pilot Association (ALPA). Subjects were paid \$12.50 per hour; the experiment lasted two days. Pilots were assigned to their respective positions in the crews - Captain or First Officer - in accordance with the position they were holding in the carrier at the time. Since data concerning pilot performance is extremely sensitive, all precautions were taken to ensure full confidentiality for all subjects. Figure 5 on the next page provides a general description of the experimental design, and outlines the experimental scenario discussed in the next section.

3.1.5. Simulated scenario

The two-pilot crews were randomly assigned to one of the three checklist conditions described above. Following one full day of training in the simulator, the crews flew a full mission simulation, comprised of four flight segments (legs), during which data were recorded. In an attempt to create a realistic simulation, the scenario included interactions with all those elements in the environment with which the pilots interact in routine flights: maintenance, ground crew, gate agents, cabin attendants, carrier dispatchers, ground control, and air traffic control (ATC). Additionally, the crews had to complete, and comply with, all the required standard paperwork, which was provided in the same format used by their airline.

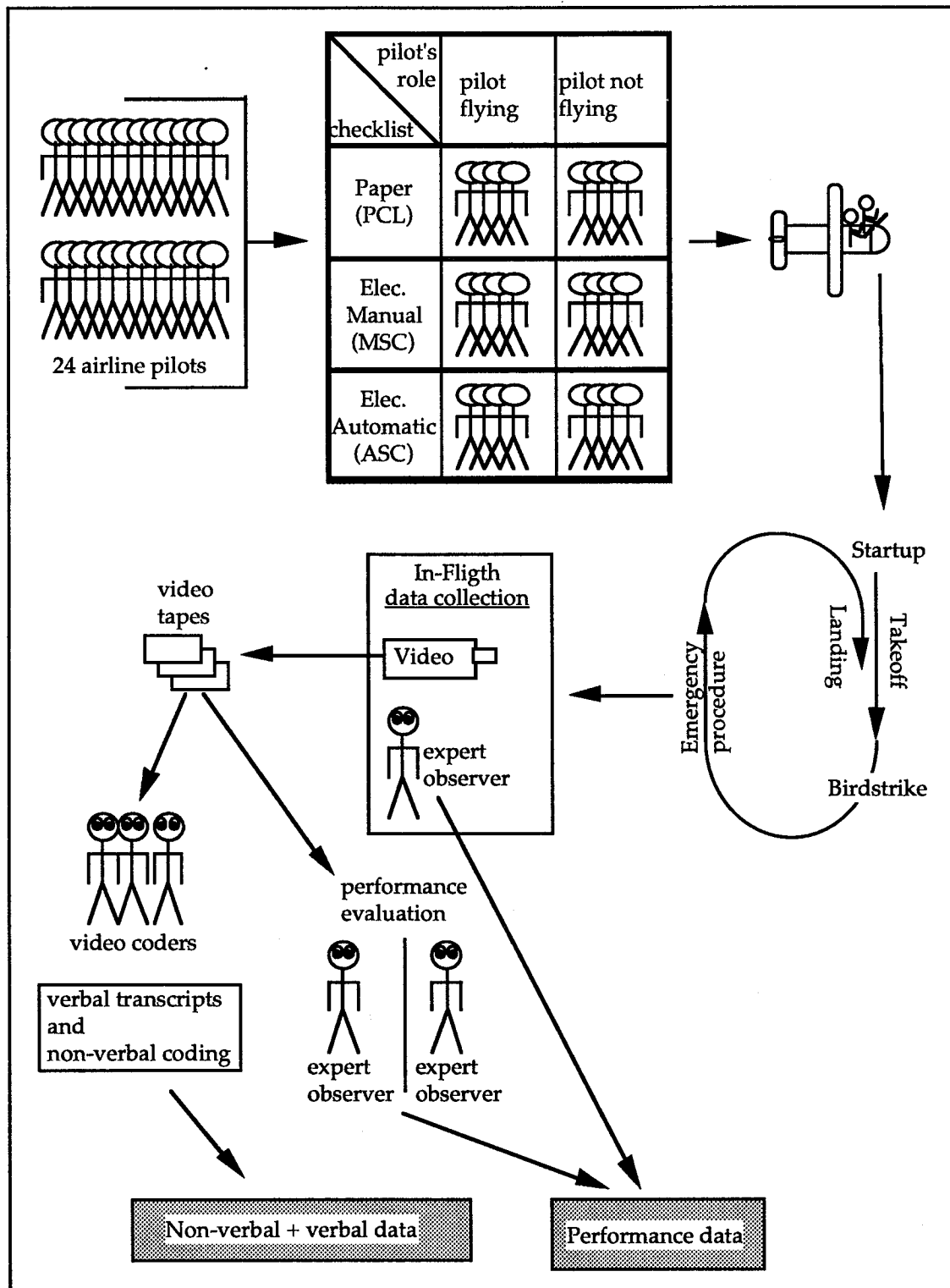


Figure 5: Experimental design

During the experimental sessions, experimenters, simulator operators and simulator engineers were situated in the experimenter operations station. Although the scenarios were all determined ahead of time and pre-programmed, data entry terminals were used to make any necessary inputs to the system. Progress of the session was monitored in real time through video, audio, and computer displays. Using different communications frequencies, the simulator operator played the role of gate agent, mechanic, dispatcher, ground crew and flight attendant; the operator even climbed up to the cockpit to deliver and pick up documents, whenever such interaction became necessary.

In order to increase scenario realism, air traffic controllers were employed to provide real-time ground-to-air interaction within the experimental environment. One controller provided clearance delivery and Automated Terminal Information (ATIS), and played the role of ground control, tower and Terminal Radar Control (TRACON). Another played the role of other aircraft at the airport and in the same airspace. A voice disguiser was used to increase the appearance of interacting with different sources within the simulated environment. The ATC operators were able to monitor the flight crew's actions and conversation via a video monitor. Thus, all radio communications from and to the a/c, as well as pseudo communications between other aircraft and ATC stations, were performed in real time, and appeared extremely realistic.

In the original experiment (Palmer & Degani, in preparation), the crews were planned to fly four legs: San Francisco to Sacramento, Sacramento to Stockton, Stockton to San Francisco, and San Francisco to Los Angeles. The last leg - San Francisco to Los Angeles - was never completed, as the simulated aircraft encountered severe engine problems immediately after takeoff, and was forced to return for an emergency landing at San Francisco. Since this leg was very demanding of the crew, and elicited highly dynamic behavior and communications, it is this particular segment of the flight on which the current research focuses.

During preparations for this final leg, the crew was advised that bird activity was sighted around the end of the runway. Following the usual checklist performance for flight and a usual takeoff, as the aircraft reached a certain target airspeed, an "Engine Fire" warning light flashed on for 9 seconds, accompanied by an aural warning of "Engine Fire." At the same time, the crew of an aircraft that took-off from a parallel runway reported to ATC that they had suffered a bird strike, and would like to return to land. While the engine fire did not persist - the warning light went out, and no other indications of fire were displayed (such as

excessively high temperatures) - both engines suffered from performance deterioration. The right engine, in particular, was affected: it produced very little thrust, with very low indicator readings. Further, a low frequency vibration sound could be heard. The pilots now had to manage the problem, using their experience and judgment, while conforming to the prescribed checklist and procedures. Specifically, they had to deal with a temporary display of engine fire warning, subsequent power loss, and a safe return for an emergency landing.

3.2. Results of preliminary analyses

In a study of the video data that focused on crew decision making, Mosier (1992) points out that although all crews did indeed succeed in making a safe landing, the decisions were not trivial. She argues that differences between the groups reflect - at least in part - the differential impact of checklist type on crew performance. While just one of the four Paper Checklist (Paper) crews shut down the left engine (an action considered as an error, since the real problem was with the right engine), five of the eight electronic checklist crews (Manual and Auto combined) made that same error. The crews that left both engines running also discussed more information concerning the status of the aircraft. It is important to stress that the highly controlled experimental scenario - which included the event-driven, on-line inputs and communication with ATC - was designed to minimize possible variations other than the particular checklist condition. Unfortunately, Mosier did not subject the findings to rigorous statistical analyses, nor look at any measure other than verbal utterances.

Performance ratings given by the expert observer who was present in the cockpit throughout all flights further support the assumption that the different checklist conditions constrained the crews to perform differently. The overall rating for communication and decision making, which was composed of six different measures, gave the Auto condition the highest performance ratings: 4.3 of 5; the Manual and Paper received 3.83 and 3.88 of 5, respectively. The observer's evaluation of workload management, which included intra-cockpit reports of workload, planning ahead for high workload situations, and usage of appropriate resources, once again rated the Auto highest (3.75 of 5), then the Manual (3.6 of 5), then the Paper (3.1 of 5). Detailed performance ratings during the flight, including performance of checklist, procedures, and aircraft handling, further highlighted the apparent effects of checklist design: Auto rated highest (4.2 of 5), Paper next (3.7 of 5), then Manual (3.6 of 5). The pattern of a clear advantage to the Auto over the two other

conditions is consistent throughout the entire set of in-flight expert performance ratings. Thorough statistical analyses of these rating data is presented in the Results section below (4.); the disagreement between the observer's ratings and Mosier's analysis is further discussed in the Discussion section (5.).

A preliminary analysis of the time required for checklist performance suggests not only qualitative, but also quantitative differences between the group. For this purpose, the duration for performance of each particular procedure was measured for each crew; in order to provide a large basis for comparison, performance times for each crew were taken from two different flight legs (Figure 6). There was essentially no difference between the groups for the time between start of engine #1 and start of engine #2; this was expected, since the interval was highly constrained, determined primarily by the time it took engine #1 to reach specific parameters. Once the crews got into the performance of other pre-takeoff checklists, however, differences in task performance times started to emerge. The Auto crews performed the "After Start" checklist significantly faster than the Paper crews ($T_{14}=3.542$; $p<.01$) and the Manual crews ($T_{14}=2.59$; $p=.02$). There was no difference between the Paper and Manual crews. The Auto crews also performed the Taxi checklist significantly faster than the Manual crews ($T_{14}=2.69$; $p=.01$), and somewhat faster than the Paper crews ($T_{14}=1.93$; $p=.07$). Again, there was no difference between the Paper and Manual crews in Taxi checklist performance time. These preliminary data confirm that, all other aspects of the task being equal, changes in checklist interface alone can impact the overall time for task performance.

Finally, in a preliminary investigation of the data, the verbal transcripts were searched for occurrences of words such as: "This, that, here," in which the auditory information alone was not sufficient to transmit the intended message. Since the information inherent in non-verbal activity was essential for comprehending these utterances, these events were called: "Action Dependent Speech" (ADS). In defining these data, the ultimate measure was the degree of relevancy which the particular pronoun presented in a manner following other research, e.g., Tracy & Moran (1983) and Haslett (1987); if there was no local relevance to the pronoun - e.g., if the utterance "I'll do *that*" did not follow a sentence that unambiguously identified what "that" specifically meant - the event was counted.

As shown in figure 7, this analysis - discussed in greater detail by the author elsewhere (Segal, 1993) - revealed that the control task and the communication task were not independent, and that pilots used information in the environment with which they

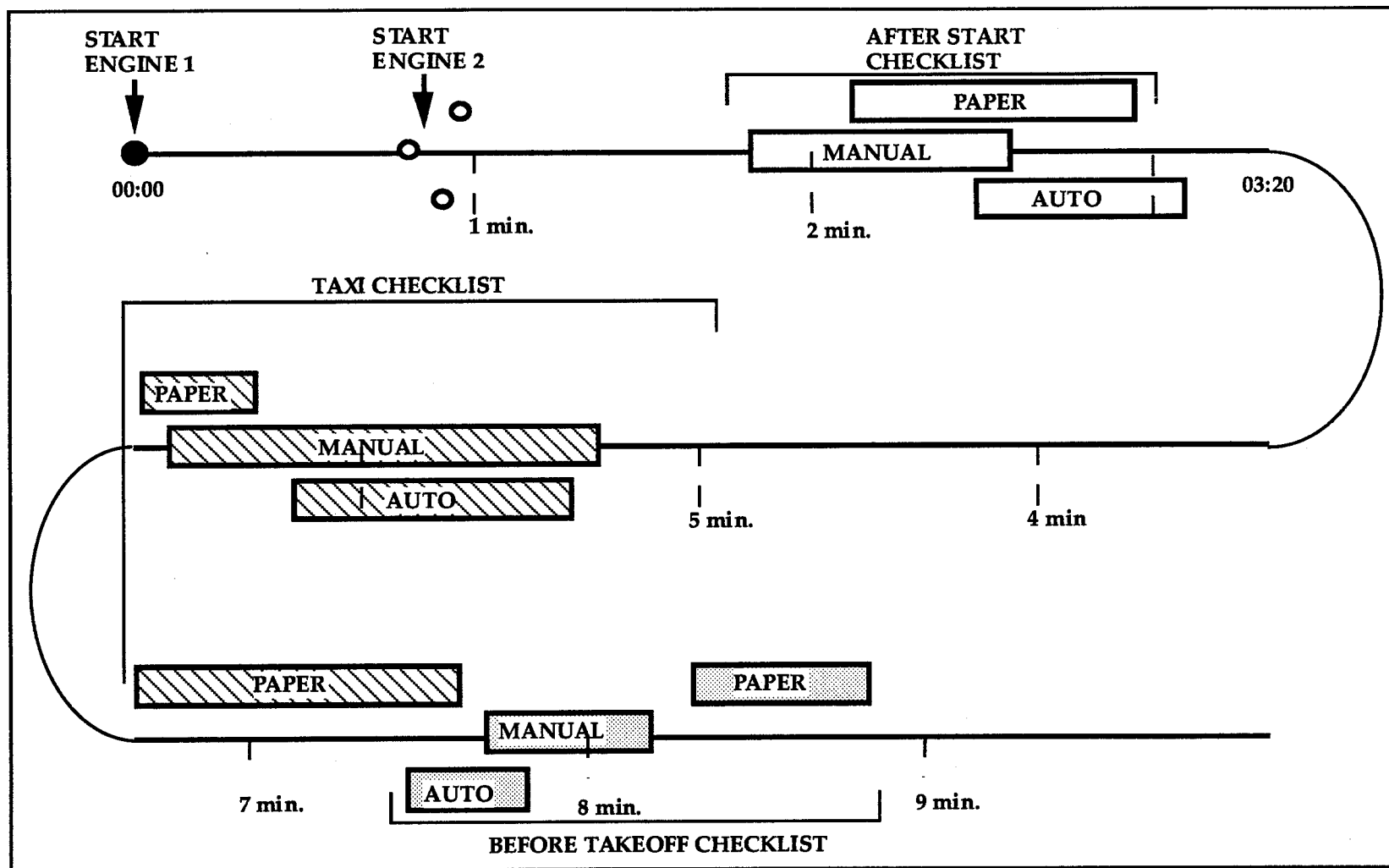


Figure 6: Checklist performance times

interacted to support their communications. Pilots in the Auto condition used more in-cockpit Action Dependent Speech than the Paper pilots, possibly relying on the added visual reference provided by the electronic checklist display. As expected, PNFs used more ADSs than PFs (Figure 7, left side). At the same time, since the visual information outside the aircraft was virtually identical for all groups, no difference between the groups was detected concerning out-the-window ADS references (Figure 7, right side).

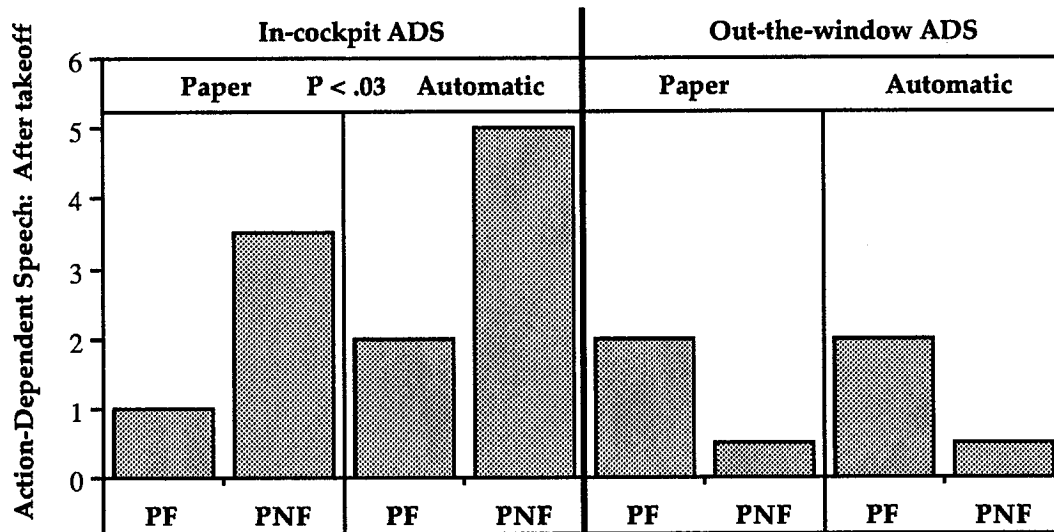


Figure 7: effect of pilot's role and checklist design on Action Dependent Speech.

In summary, a systematic pattern of difference between the groups has been suggested by various preliminary analyses: the differences in overall task performance between the different conditions noted by Mosier (1992), the differences in the time it took crews to perform the checklists, the in-flight observer's ratings of individual and crew performance, and the preliminary analysis of Action Dependent Speech described above. Further, low variability in the pilots' level of competency may be assumed - all were trained by and fly for the same airline, and received similar training in the simulator - and thus it may safely be argued that the manipulation of checklist design indeed created task environments that resulted in significantly different performance. The research presented below further explored the dynamics that took place in the cockpit, focusing analysis on the interdependence between the verbal and non-verbal information that emerged from the crew's interaction with the different checklist designs within this highly demanding experimental scenario.

3.3. Hypotheses

Based on the discussion above, the following hypotheses were proposed:

- 1.a. The design of the electronic checklists (both Auto and Manual) - which are based on a touch-sensitive CRT interface - affords the emergence of non-verbal information, in the form of touching the touch-sensitive checklist displays. This will encourage crews that use it to rely on more non-verbal communication cues than those crews using the paper checklist design (Paper).
- 1.b. Between the two electronic designs, the automated Auto provides fewer opportunities for non-verbal communication than the Manual, and will thus result in less reliance on non-verbal (consequential) communication.
2. The two different roles in the crew - pilot flying (PF) and pilot-not-flying (PNF) - will yield different reliance on non-verbal communication, with the PF - who was busy flying the plane, and thus did not perform the checklists - doing more "detection" of PNF's in-cockpit activities, and the PNF - who was in charge of completing the relevant checklists - doing more "emission" of consequential information.
3. Differences in checklist design will result in significant differences in crew performance, as measured by performance ratings provided by the expert observers.

3.4. Proposed data interpretation and coding

One of the primary reasons that non-verbal behavior has been left untouched is the methodological difficulty which underlies all studies of non-verbal communication. This difficulty was well illustrated in the study by Chapanis et al. (1972), described in section 1.2.3. above. Earlier experimental studies of multiple-operator task performance seem to have encountered similar difficulties (Smith & Wilson, 1963; Wiener, 1963; Pollack & Madans, 1964). In the preparation of this proposal, the author explored the different methodologies available for transcription and analysis of non-verbal behavior. As an example, the hand movement code proposed by Friesen et al. (1979) was considered, where a hand act was defined as movements in the hand which could be coded as either illustrator, manipulator, or emblem. They applied the code to videotapes of conversations and inspected the reliability of their proposed methodology, with results that showed a high degree of intercoder agreement. Ekman & Friesen (1976) developed a procedure for measuring visibly different facial movements, where any facial movement (observed in photographs, motion picture film, or videotape) could be described in terms of anatomically

based action units. The anthropologist Ray Birdwhistell developed several methodologies for encoding and analyzing facial gestures and body movements (1970). These studies focus on such things as facial expressions and body posture as they manifest themselves in the process of direct interaction *not in the context of crew-machine interactions*, in which control activities and consequential communication play primary communicative roles. Thus, following the theoretical background presented in the introductory discussion, and in accordance with the hypotheses described above, this research was designed on the basis of a different approach to the interpretation of the video and audio data collected in the simulator.

Obviously, a key issue was the initial definition of "communication" - i.e. when was a message present, was it transmitted or emitted, was it received or perceived, was it understood? These questions seem crucial in the process of constructing a methodology for coding and interpreting the recorded video and audio data. One useful source for the analysis of non-verbal communication is the field of animal behavior, where scientists have focused much effort into the analysis of animal communication systems. As a guiding rule, the following principal seems to be invaluable, and especially relevant to the current paradigm: the meaning of the communication is the response it elicits. In this sense, the information present in the activity of one organism can be considered as "communication" only if it elicits a response from another organism (Alcock, 1989). While this definition does not account for messages that are received without immediate, explicit confirmation, or messages that demand that the receiver not respond explicitly, it is assumed that in the context of the temporal constraints imposed by the crew's need to perform the flight task, these kinds of events were minimized. Accordingly, the coding scheme was designed to capture those instances where one pilot's reaching for a certain display elicited a look at that display from the other pilot.

In order to facilitate the analysis of audio recordings, the entire flight segment was transcribed verbatim; this, as well as the critical task of video coding, was accomplished with the help of research assistants employed by NASA-Ames through the San Jose State University Foundation. All verbal activities were transcribed; the coders parsed the speech into naturally occurring "speech acts," as defined by Levinson (1983): "The making of a statement, offer, promise, etc. in uttering a sentence, by virtue of the conventional force associated with it." Particular emphasis was put on identifying the intended receiver for each speech act; following the categorization of speech acts by recipient, the current analysis specifically examined those speech acts that were intended as in-cockpit, intra-crew,

communications. Thus, the analysis of verbal communication focused on speech interactions between the pilots, and did not include verbal interactions between the pilots and other elements in the task environment, e.g., ground stations, traffic control centers, flight operations. While it is unquestionable that such speech activity provides much consequential information, this study did not look at this particular form of crew interaction.

Next, the coders looked at the non-verbal activity in the cockpit. Four categories of activity were defined, and the occurrence and duration of each was entered into a time line which included the activities of both pilots in each crew. These categories of observable, non-verbal activity were:

1. "Look": Pilot looks at his/her pilot's checklist display
2. "Touch": Pilot touches/manipulates own checklist display;
3. "Point-In": Pilot points at own checklist display
4. "Point-out": Pilot points out the window

Once the video coding process began, the data collected by the three coders for a particular video tape was compared. It became clear that the coding task could not be performed reliably by any one coder, and that if the task were performed by independent coders, the correlation between their data files would not satisfy basic requirements. It was decided, therefore, that three coders would work interactively on all the video data, consulting with each other to clarify ambiguous segments, and coding on the basis of consensus. For the purpose of performing time-series analysis, the data were subsequently entered into MacShapa (Sanderson, 1993), a unique computer program designed for Exploratory Sequential Data Analysis, described in greater detail in the next section.

As is the case in most studies of communication, in which the process of measurement relies on humans, rather than machines, the statistical analysis was performed on both quantitative and qualitative data, i.e. nominal, categorical, data (see Kennedy, 1983, for a discussion of data types in general, and nominal data in particular). In the measurement of non-verbal activity, the data consisted of duration of action, e.g., time spent manipulating the display. The primary reason for choosing time of activity - rather than, for example, number of actions - was that no clear semantic scheme was found for labeling activity, and thus the count of individual actions seemed impossible. For example, in performing a ten-item checklist procedure, some pilots reached for the display five times and touched it, withdrawing their hand for a brief instant between touches, while others left their hand "hovering" over the surface of the display throughout the entire checklist, making slight input movements which were virtually undetectable on the video records. It was thus

decided to measure the time spent interacting with the display - i.e., the duration of time a pilot spent engaged in display manipulation - rather than the particular number of times the display was touched. The verbal data were transcribed and categorized, and analysis involving these data examined the frequency of occurrence of intra-crew, in-cockpit, Speech Acts (Levinson, 1983).

In order to better understand the impact of checklist design on different task scenarios, five key segments were identified along the leg. The first was performance of the After Start checklist; the second, the period from the end of that checklist, through taxi, and to the start of the Before Takeoff checklist; the third, from that point to initiation of the takeoff role; the fourth, from takeoff to the point at which the birdstrike occurred; and finally, the segment of flight from birdstrike to landing, including performance of the emergency procedure. Obviously, the different segments varied in their length; for the purpose of analysis, wherever appropriate, corrections were made for differences in segment times. Although the analysis included all phases, the following discussion will focus primarily on the comparison of the first phase - After Start checklist performance - and the last phase - the emergency procedure. These two particular phases were chosen because they represent two very different flight scenarios: the After Start checklist is a well practiced and highly structured procedure; the emergency procedure was unexpected, and, by virtue of the ambiguous decision problems it posed, generate much spontaneous crew interaction.

Additionally, throughout the planning of this proposal, it seemed essential to include the analysis of video recordings by experienced pilots to evaluate and rate pilot and crew performance. The introduction of human operators into the measurement and analysis process seems to be an essential step, similar to the use of native speakers for the analysis of speech communication. The importance of using trained expert observers is confirmed by Chidester et al. (1989), who discuss the use of experts to detect and recognize errors both in real-time in the cockpit, as well as after the flight in reviewing videotape records of cockpit activity. As part of the original study (Palmer & Degani, in preparation), the video tapes were studied independently by two observers - former airline captains - both of whom were selected by NASA-Ames ASHFRD as domain experts. They worked independently, and focused their analysis on performance errors. Their analysis of crew performance in general, and of performance of the checklist in particular - as well as the ratings provided by the in-flight observer - will be presented in the Results section. Figure 8 describes the process of data collection, coding and transcription, and summarizes the dependent variables upon which the analysis was based.

Several abbreviations will be used throughout the following sections. The term "SA" represents "speech-act," specifically, intra-crew verbal interactions. The term "NV" will refer to those non-verbal activities that were coded off the video tapes by the trained coders. Unless otherwise specified, these will include all four categories of activity: pilots *looking* across the cockpit at their crewmembers' display, reaching to manipulate and *touching* their own display, *pointing in* at the display and *pointing out* at conflicting traffic and landmarks. "PF" will refer to the pilot flying the plane, and "PNF," to the pilot who was not flying and was in charge of performing the checklist and other system-related tasks.

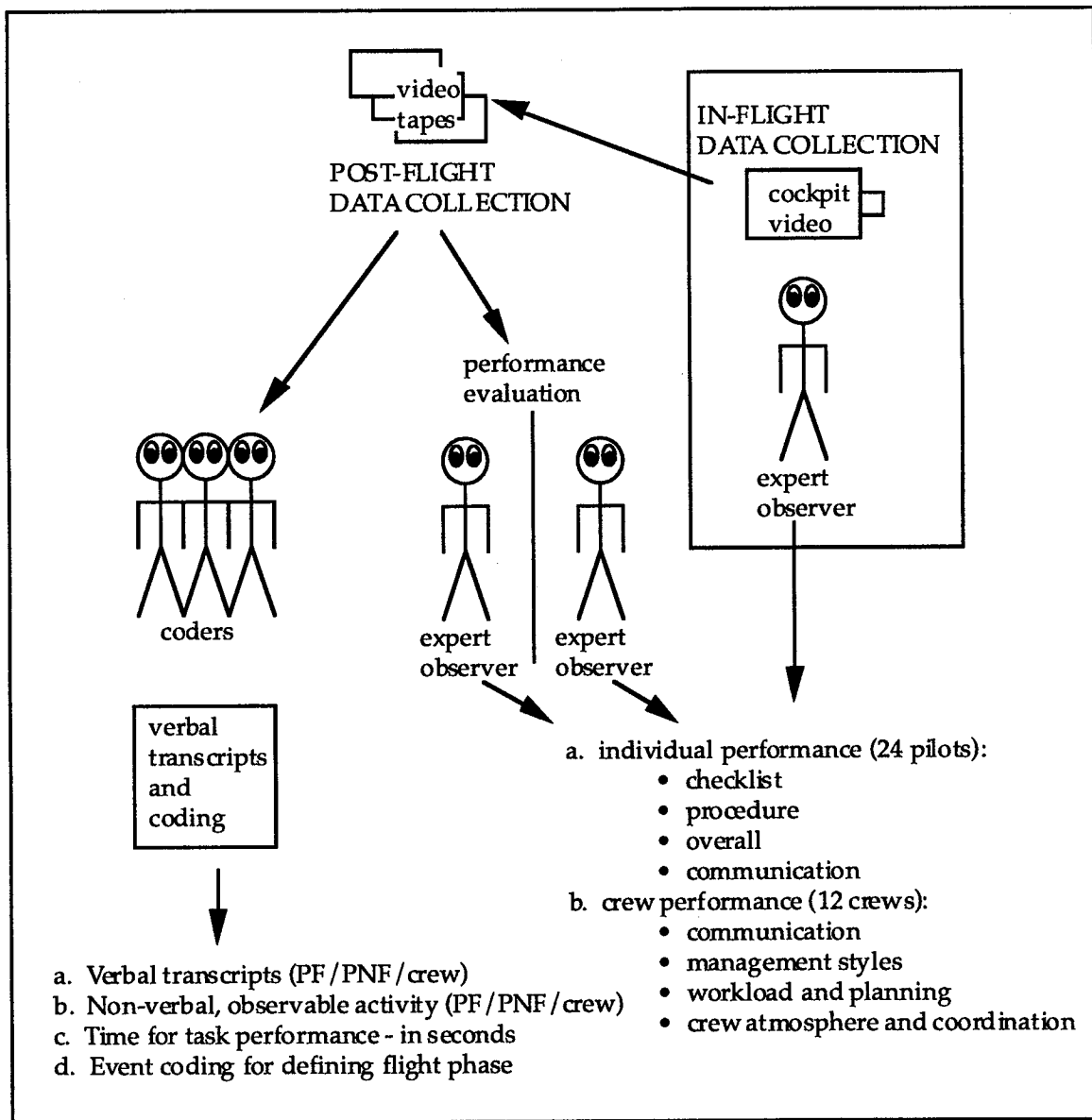


Figure 8: Dependent variables

4. Results

4.1. Non-verbal activity

This section begins with a report of the findings concerning effects of checklist design (Paper, Manual and Auto), pilot's role (PF vs. PNF) and flight phase on the non-verbal activity observed in the cockpit. The results first consider differences in overall non-verbal activity, then focus on two mutually exclusive, yet dependent, categories of activity: manipulation of the checklist display by one pilot, and looking at that display by the other pilot. Note that in order to correct for differences in duration of flight, the analysis focuses on the proportion of total flight time during which activity was observed in the cockpit, i.e., the units describe "observed activity time divided by total flight time."

- *Effect of checklist design on non-verbal activities:* An analysis of the effect of checklist condition on total time of non-verbal cockpit activity - normalized by dividing that time by the total flight time - yielded a significant difference between the three groups ($F_{2,23}=5.036$; $p<.02$), as shown in Figure 9. A pairwise comparison using the Tukey test yielded a significant difference (at $p=.05$) between the Paper condition and the Manual condition, with pilots in the Paper condition performing fewer observable actions. The two electronic conditions of Manual and Auto did not differ significantly ($p>0.1$).

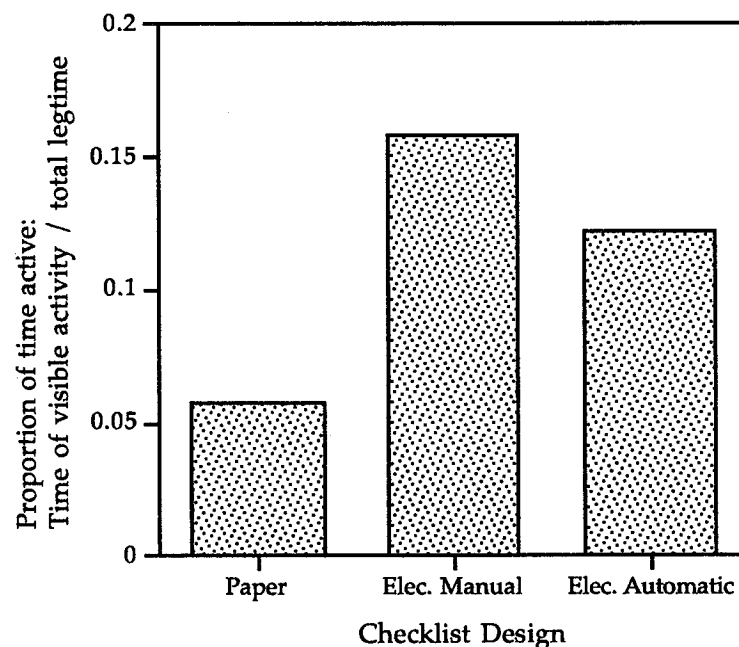


Figure 9: Effect of checklist design on non-verbal activity

Since three of the four NV categories were specifically related to the checklist display, and since the checklist interface located at the bottom half of the touch-sensitive CRT display was interactive only for the two Electronic groups - it is not surprising that crews in these groups were seen to be more active than crews in the Paper group. The lack of a significant difference between the Manual and Auto group is surprising, given that the automation was designed to assume a large portion of the pilots' checklist activities.

- *Effects of checklist design and pilot's role on display manipulation:* A two-way ANOVA was used to test the effect of condition and pilot's role on the proportion of time at which pilots manipulated the checklist display, using $[(\text{TouchTime} + \text{PointTime}) / \text{Legtime}]$ as the dependent measure; to reflect the interdependence of crewmembers, PF and PNF were blocked by crew. As can be seen in Figure 10, there was a main effect for both condition ($F_{2,9}=5.43, p<.01$) and role ($F_{1,9}=75.98; p<.001$). A significant interaction effect was also observed ($F_{2,9}=9.25; p<.01$). The PNFs spent more time than their crew-members manipulating the checklist display; this finding is not surprising, given that their role in the crew specified that they be the ones performing the checklist. Similarly expected was the effect of condition on the PNF's activities, in which the PNFs using the Manual checklist interacted with their display more than the two other groups: more than the Paper group due to the interface, and more than the Auto group due to lack of automation.

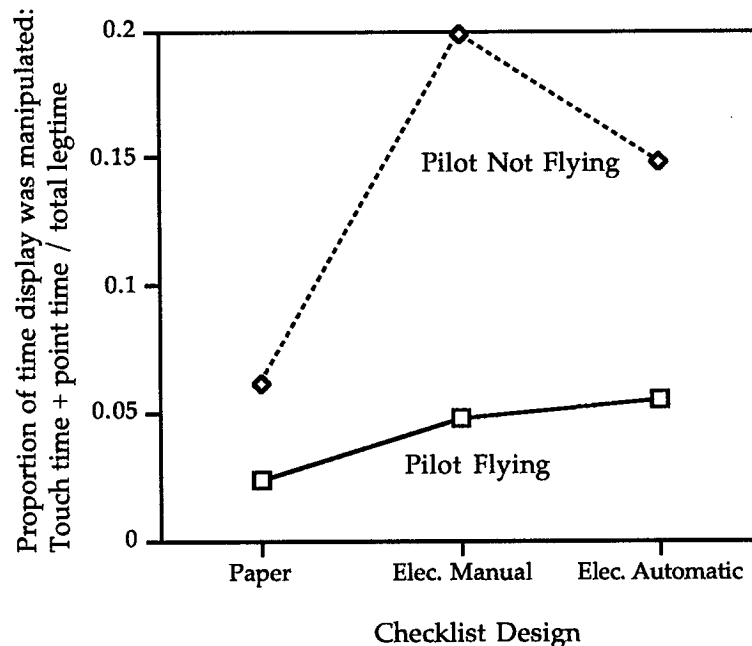


Figure 10: Effects of checklist design and pilot's role on display manipulation

- *Effect of checklist design on PNF's display manipulation:* Following the above finding, a more focused comparison of the PNF's display manipulation times between the three groups (dashed line in Figure 10) yielded a significant effect ($F_{2,11}=12.98$; $p<.01$). Further pairwise comparisons using the Tukey test yielded a significant (at $p=.01$) difference between the Paper and the two other groups, but no difference between the two Electronic groups. This finding also confirms what is evident from Figure 10, that is, that the difference in activity between the three conditions detected in the data described in Figure 9 is attributable primarily to the difference in activities of the PNFs.

- *Effect of checklist design and pilot's role on monitoring other's display:* In order to determine the effect of checklist interface and task on the proportion of time at which each pilot looked across the cockpit at the other pilot's display, a two way ANOVA (Condition x Role) was performed on LookTime/Legtime (Figure 11); PF and PNF were blocked by crew. While no main effect was observed, the interaction between the two factors was marginally significant ($F_{2,9}=3.14$; $p=.09$).

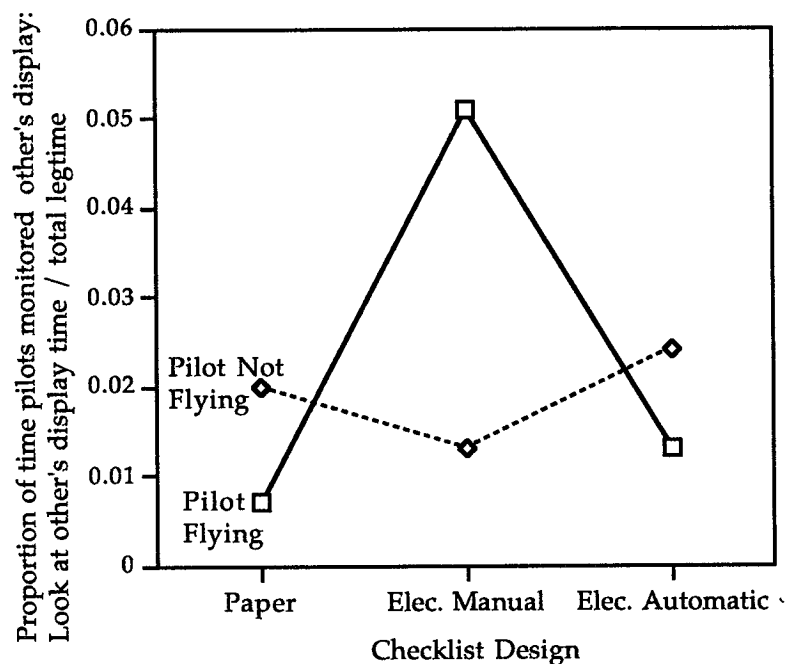


Figure 11: Effect of checklist design and pilot's role on looking at other's display

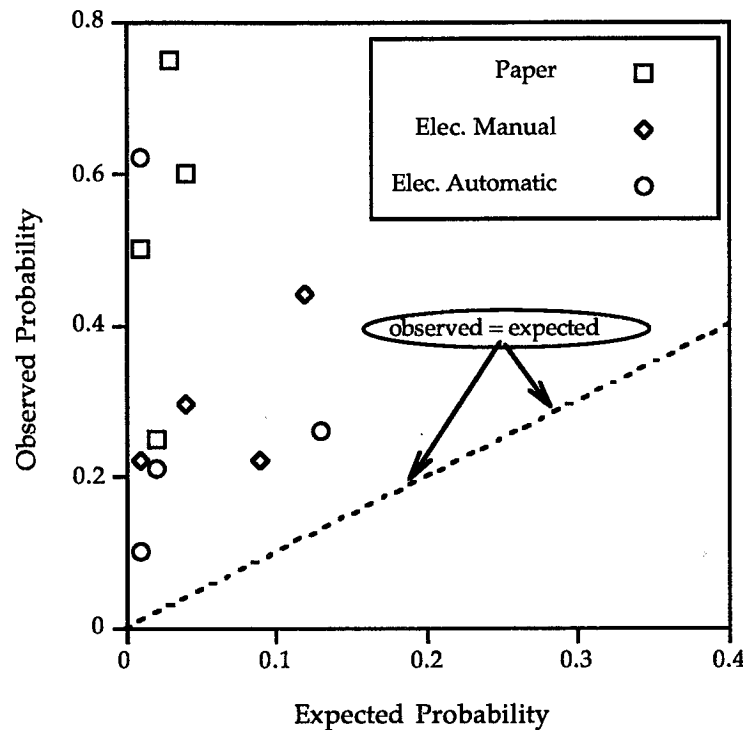
Crews in the Manual condition seemed to be the main reason for the interaction, with PFs in this condition looking much more than PFs in the two other groups, while PNFs looked less. Given the above findings (Figure 10) suggesting that the PNFs in the Manual group were

busier manipulating their display than the PNFs from the two other groups, it may be reasonable to speculate that while the Manual PNFs' workload - greater than the other groups because of the absence of automation - did not allow them to look at their crew member's display as much, the activity they displayed lead their PFs to look at them more than PFs from other conditions. Interestingly, the PNFs from the Paper and Auto groups looked across the cockpit more than their PFs. An ANOVA testing the effect of role for these two conditions yielded a significantly larger amount of "looks" for PNFs ($F_{1,14}=5.208$; $p=.04$).

- *Effect of condition on PF's looking activity:* Focusing the analysis of the "look" behavior on the PFs (solid line, Figure 11), an ANOVA testing the effect of condition on look time yielded a significant difference between the groups ($F_{2,9}=5.692$; $p<.05$). A pairwise comparison using the Tukey test identified a significant difference (at $p=.05$) between the Manual group and the two other groups, but not between the Paper and Automatic group. Following the argument proposed above, this finding makes sense: since there was more activity displayed by the PNFs of the Manual group, there was more information for the PFs of that group to gain from looking across the cockpit at their crewmember's display.

- *The probability of transition from one's "touch" to another's "look":* The relationship between expected and observed probabilities of observing a "look" following a "touch" is presented in Figure 12. In order to test whether the non-verbal behavior of touching one's own display may predict a cross-cockpit look from the other pilot, MacShapa (Sanderson, 1993) was used to perform a lag sequential analysis on the entire non-verbal data stream. This analysis focused on whether a pilot tended to look at their crewmember's display immediately after that crewmember manipulated that display; thus, the analysis focused on events that follow each other immediately, without any intervening events.

Of the twelve crews which made up the three groups, 9 crews - three of four from each experimental group - yielded Markov Z scores that suggest that the observed probability of "look" following "touch" was significantly higher (at the $p=.05$ level) than expected by chance (see Faraone and Dorfman, 1987, for a description of the Markov Z statistic). While the nature of lag sequential analysis requires that caution be used in drawing conclusions about actual sequences (Gottman & Roy, 1990), this finding further confirmed the close connection between activity by one pilot and the monitoring of that activity by the other.



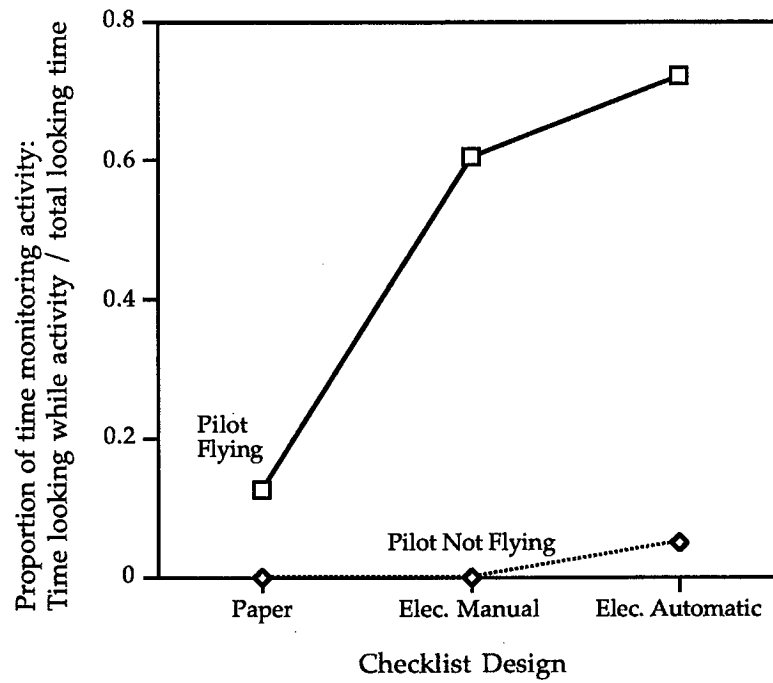


Figure 13: Effect of checklist design and pilot's role on the ratio of looking while the other is manipulating to overall looking

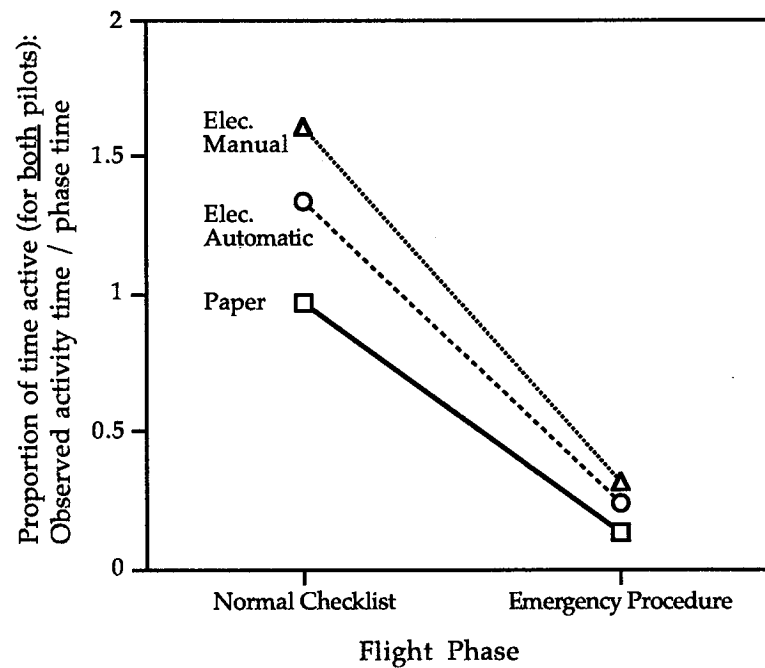


Figure 14: Variations in non-verbal activity across two flight phases

- *Differences across flight phases:* To examine the effect of the different flight phases on non-verbal activity, a two-way, repeated measures, ANOVA (Condition x Phase) was conducted, looking at the time of non-verbal activity (corrected for differences in phase times). Recall that this contrast compared the highly routinized After Start checklist phase with the unexpected emergency phase which followed the birdstrike (see Figure 14). The ANOVA yielded a main effect for condition ($F_{2,9}=4.69$; $p<.05$) and phase ($F_{1,9}=72.61$; $p<.001$). There was no interaction between the variables.

Note that since activity for both crewmembers was coded separately, then summed up for each crew, the measured ratio may indeed exceed 1, as it did for the two electronic groups in the early phase of the flight. The pattern across phases is not surprising, since the After Start checklist phase was relatively short and rigidly structured to include much activity, while the emergency phase took much longer, and included the return to landing segment in which not much was performed. The primary reason for presenting these data is for later discussion of effects of flight phase.

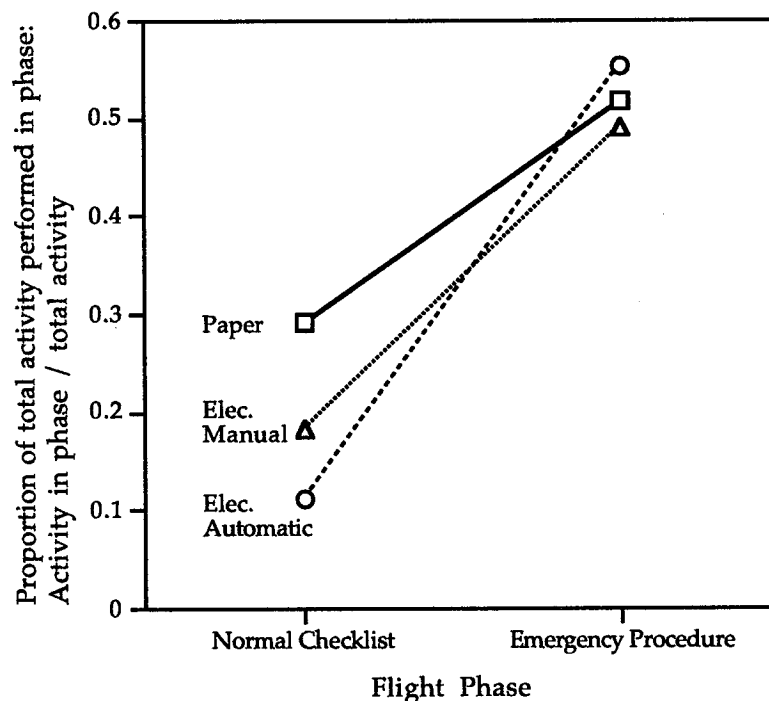


Figure 15: Distribution of activity across different flight phases

- *Effect of checklist design and flight phase on distribution of non-verbals* A two-way, repeated measures, ANOVA testing the effect of condition and phase on the proportion of total non-verbal activity performed in each phase (NV-in-phase/total NV) yielded a main effect for

condition ($F_{2,9}=5.54$; $p<.05$) and phase ($F_{1,9}=186.22$; $p<.001$); there was a significant interaction between condition and phase ($F_{2,9}=7.28$; $p<.02$). As is shown in Figure 15, while all three groups performed a large proportion of their activity in the emergency phase of the flight, the two Electronic groups seem to display a relatively larger proportion of activity in that phase. This difference is most noticeable for the Auto, which goes from lowest portion of activity in the earlier phase to the highest in the later phase. A similar trend can be seen in the Manual data; the Paper group displayed the most moderate slope of all three.

Reviewing the above findings concerning non-verbal activity, several patterns seem to emerge. Overall, the Paper group tended to display less cockpit activity than the two Electronic groups; within the Electronic conditions, the Manual crews tended to engage in more control and monitoring activity than the Automatic crews. For all three groups, the pilots' roles in the crew had a significant impact on the type and quantity of activity, with PNFs tending to engage in more control inputs while PFs engage in more monitoring behavior. The connection between these two activities - which, by virtue of being distributed between the two pilots, may provide insight to their cooperation strategies - proved to be quite strong, as indicated by the high probability of observing a transition from one's touch to another's look, as well as by the large ratio of concurrent looking to total looking. Here, in particular, the data show an increased ratio for both Electronic groups, suggesting that the added amount of non-verbal activity may be supporting the pilots' reliance on monitoring and consequential communication.

It is important to note that since the analysis focused only on pilots' looks at the checklist display, the data probably represents the lower measure of the true number of cross-cockpit looks which are occurring. Indeed, throughout the video coding sessions, pilots were seen to look at other pilots interacting with systems which were located in different places in the cockpit, e.g., on the overhead panel, or the pedestal console between the two pilots. Given the large range of locations of activity, and, consequently, the large range of possible looks across the cockpit, and since the only difference in cockpit design between the three conditions was in the functionality of the checklist display, all control activity and cross-cockpit looks that were not directly related to that system were not coded, and were not included in the analysis.

Finally, the difference in distribution of activity over different flight phases indicates that while automation supported reduced activity under normal flight conditions, it elicited a relatively higher proportion of activity under emergency flight conditions.

4.2. Verbal utterances

In this section, a brief analysis of verbal communication first looks for fundamental differences between the three experimental groups, then at the effects that different flight phases exerted on these groups. A subsequent analysis of in-cockpit verbal data focused primarily on the relationship between verbal and non-verbal activity, the findings of which are discussed in the next section.

The verbal transcripts were used to measure both frequency and duration of in-cockpit Speech Acts. Preliminary data analysis tested the comparative value of using one as opposed to the other; when virtually identical results were yielded for both, it was decided that the measure of frequency of Speech Acts be used consistently throughout all analyses involving verbal transcripts.

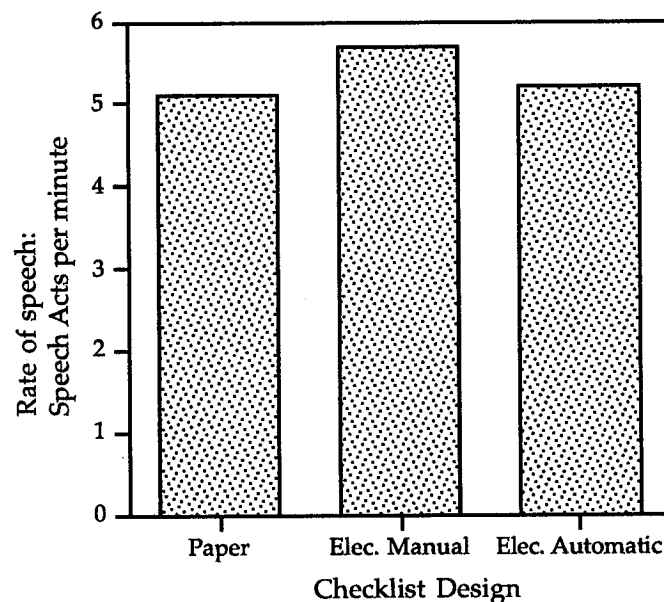


Figure 16: Effects of checklist design on rate of speech

- *Effect of checklist design on number of Speech Acts:* The three groups did not differ in the rates of verbal communication within the cockpit (Figure 16). An ANOVA performed on the number of in-crew communication speech acts, adjusted for the variance in the length of flight time (Speech Acts / leg time) yielded no significant difference between groups ($F_{2,9} = 0.321$; $p = .73$).

• *Effects of condition and flight phase on distribution of Speech Acts:* Since no difference was apparent in overall speech rate, it was interesting to see whether the different groups distributed their speech throughout the flight in a similar pattern. A two-way, repeated measures, ANOVA (Condition x Phase) was performed, using (speech-acts in phase / total speech-acts) as the dependent measure (see Figure 17). A main effect for phase was found ($F_{1,9}=409.01$; $p<.001$), as well as a Condition x Phase interaction effect ($F_{2,9}=6.76$; $p<.02$). The main effect was expected, since the emergency procedure phase lasted considerably longer than the initial after-start checklist phase. The interaction effect suggests that when the task was well defined, the Auto allowed performance of the task with minimal verbal interaction; in the emergency procedure, however, crews in the Auto condition performed the task using a relatively higher amount of in-cockpit communications, suggesting a less balanced distribution of speech across the flight.

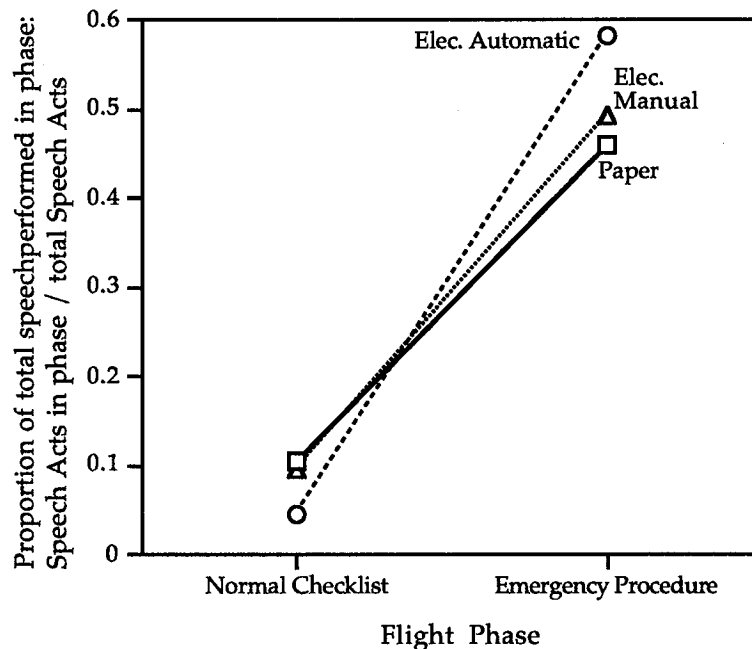


Figure 17: Proportion of Speech Acts uttered in two different flight phases

4.3. Relationship between verbal and non-verbal

• *Effects of checklist design and role on NV/SA:* In order to determine the relative occurrence of non-verbal and verbal activity, the total time of observable non-verbals was divided by the number of in-crew verbal utterances. A two-way ANOVA, in which pilots were blocked by crew, testing for the effect of condition and role on the ratio of NV/SA yielded a significant main effect for condition ($F_{2,9}=5.05$; $p<.05$) and for role ($F_{1,9}=258.12$ $p<.001$). There was also an interaction effect ($F_{2,9}=9.95$; $p<.001$). Overall, as shown in Figure 18, the PNFs performed more non-verbal activity in relation to their speech than did the PFs. For both crew members, the crews in the Electronic condition seemed to have a larger ratio of non-verbals to verbals than did the Paper crews, with the Manual exhibiting the highest action/speech-act ratio. It was established earlier that the groups differed in NV activity (Figure 9), and that they did not differ in SA activities (Figure 16), hence the relationship described here may seem obvious. Nevertheless, it was important to establish the difference between the groups along this fundamental measure.

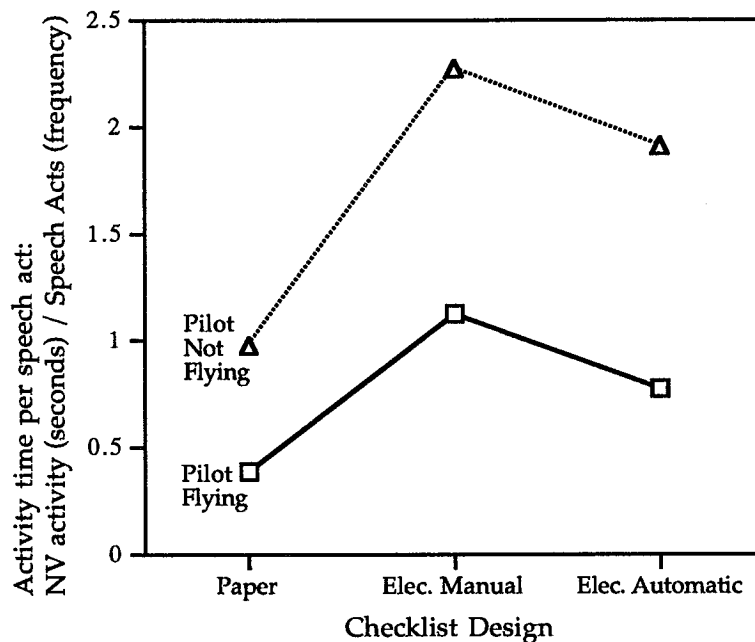


Figure 18: Effect of checklist design and pilot's role on ratio of non-verbal activity and Speech Acts

This analysis emphasizes the point that the Electronic interface created a task environment in which both pilots were more active than the Paper pilots without a concurrent increase in

the amount of in-cockpit verbal communication. Since the analysis of cross-cockpit looks validated that PFs indeed looked at these actions, the Electronic checklist can be seen as providing added information in the form of consequential communication.

- *Effect of checklist design and flight phase on NV/SA:* A two-way, repeated measures, ANOVA testing the effect of condition and phase on NV/SA yielded a main effect for condition ($F_{2,9}=5.99$; $p<.03$) and phase ($F_{1,9}=53.26$; $p<.001$), and no interaction effect (see Figure 19). In general, the ratio of NV/SA was higher in the normal checklist phase than in the emergency checklist phase; in comparing the three different design conditions, the two Electronic groups exhibited more activity relative to speech acts than the Paper group. These data should be considered in the context of the data presented in Figure 17 above, which suggested that the two Auto crew members performed a large proportion of their total speech acts in the emergency condition. While the Paper and Manual crews were virtually identical in the proportional distribution of speech acts across the two phases (Figure 15), the ratio of non-verbal to verbal yielded a greater similarity between the two Electronic groups, with the Paper group scoring lower than both, across both phases of flight.

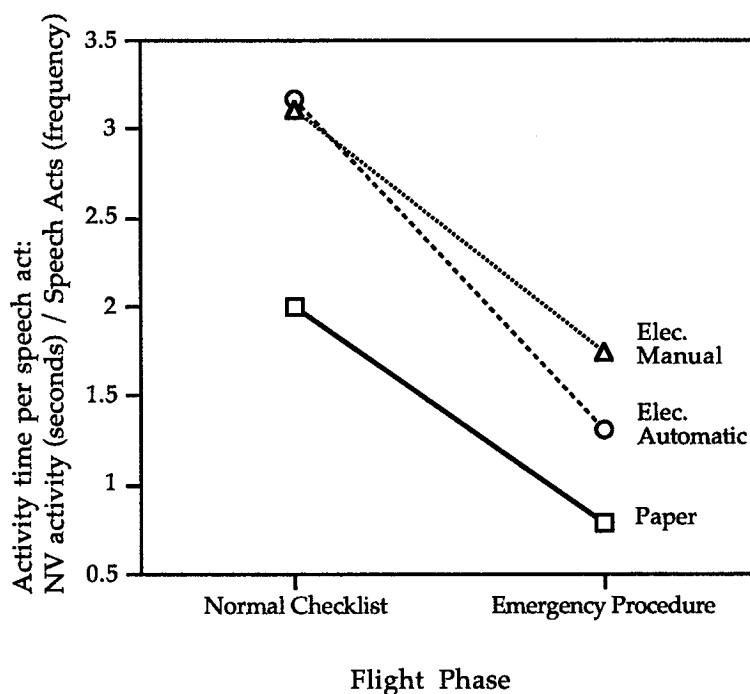


Figure 19: Variations in activity / speech ratio across different flight phases

In summary, it seems that the overall rate of speech in the cockpit was affected neither by checklist design, nor by the pilots' different roles within the crews. This finding alone is interesting, since automation of the checklist was designed to alleviate some of the workload on the crew, and presumably, could have made some intra-crew speech redundant. Since the crews did vary in the amount of non-verbal activity in the cockpit, it is interesting that the increase in activity was not accompanied by an equal increase - or decrease - in speech. One possible interpretation is that the information inherent in activity provided the Electronic crews with sufficient communication material to make added speech unnecessary.

4.4. Performance

As was described in the Methodology (section 3), crew performance was measured in three different manners: 1) During the flight, an in flight observer - who had accompanied all crews throughout all their missions - rated the crews for performance; 2) Following the flights, two independent observers went over the video recordings, and rated each crew and crew member for performance; 3) A global performance measure relating to performance of the emergency checklist - and, specifically, to whether the crew shut-down the wrong engine - was used by Mosier (1992) in her analysis of crew performance. As will become evident from the following discussion, there was no agreement between the different raters regarding the performance of the different crews. This seems due primarily to the structure of the performance rating scales defined by the original investigators (Palmer and Degani, in preparation).

- *Performance ratings by observer:* The three checklist groups - Paper, Manual and Auto - were rated for performance of different aspects of crew cooperation by the in-flight observer (see Figure 20). A one-way ANOVA testing the effect of condition on these ratings yielded a marginal difference ($F_{2,47}=2.463$; $p=.097$). Subsequent pairwise comparisons yielded a significant difference (at $p<.05$) between the Paper and Auto crews ($F_{1,31}=4.1$; $p=.05$). The Manual crews did not differ significantly from either. The mere fact that there was a trend for difference between the three conditions is interesting, since the particular measures discussed here addressed "crew communication," "management style" and "coordination," variables which theoretically should not have been affected by level of automation.

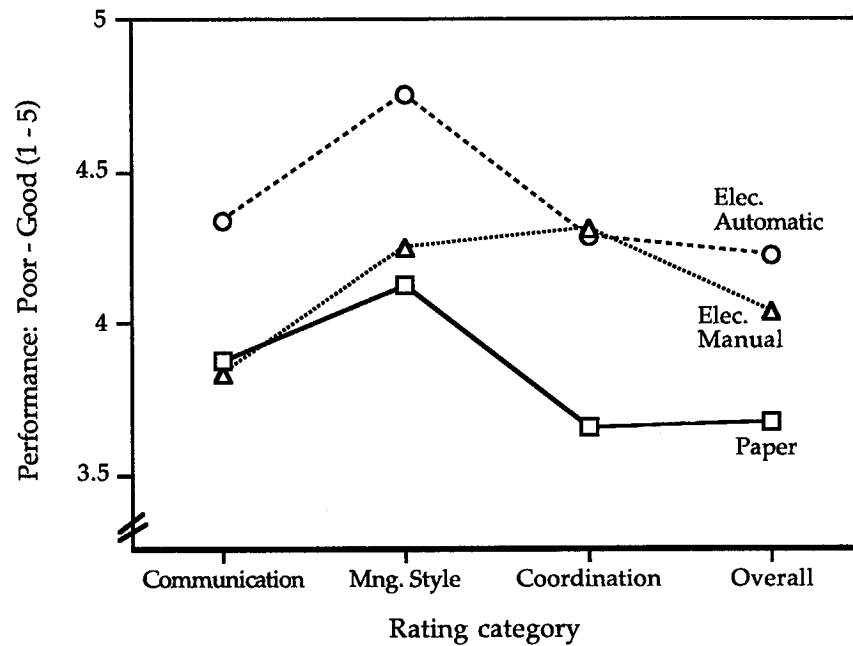


Figure 20: Performance ratings by in-flight expert observer

- Correlations:** In order to test the degree of relationship between performance and transcribed variables for all 12 crews, a correlation matrix was constructed, and coefficients calculated (Table 2). Note that although the above analysis of NVs indicates a significant difference in performance between the groups, no relationship was found between these measures of activity and performance measures. At the same time, while the three groups showed no significant difference in the rate of Speech Acts within the cockpit, these were found to be negatively correlated with performance measures. It seems that, at least from the perspective of the in-flight observer, crews that performed the task with less speech rated higher on measures such as communication, interpersonal style and crew atmosphere and coordination.

<u>Variable</u>	<u>NV/legtime</u>	<u>SA/legtime</u>	<u>NV/SA</u>
Overall performance	-.011	-.632*	.339
Communication	-.008	-.507	.276
Interpersonal style	-.064	-.709**	.317
Atmosphere and coordination	-.026	-.652*	.33

[Note: Based on n=12. * $p < .05$; ** $p < .01$]

Table 2: Pearson correlations of performance measures and transcribed NVs and SAs

Since another source for performance ratings focused on crew's performance in the emergency phase of flight (see discussion of Mosier's study below), a detailed analysis of the in-flight observer's ratings focused on that particular high-workload phase of flight. The in-flight observer had rated individual crewmembers for performance during each phase of the flight, again, rating their performance in a set of prescribed categories along a scale of 1 to 5. Figure 21 presents the data pertaining to three categories of behavior: Situational Awareness, Workload Distribution, and Communication and Coordination. ANOVAs comparing the ratings for the three groups yielded significant differences for the ratings of Situational Awareness ($F_{2,23}=4.433$; $p=0.02$) and Workload Distribution ($F_{2,22}=3.874$; $p=0.04$); there was a marginal difference in ratings of Communication & Coordination ($F_{2,23}=2.579$; $p=0.1$). Overall, the Manual group rated lower than either Paper or Auto; pairwise comparisons did not yield a significant advantage for the Auto over the Paper.

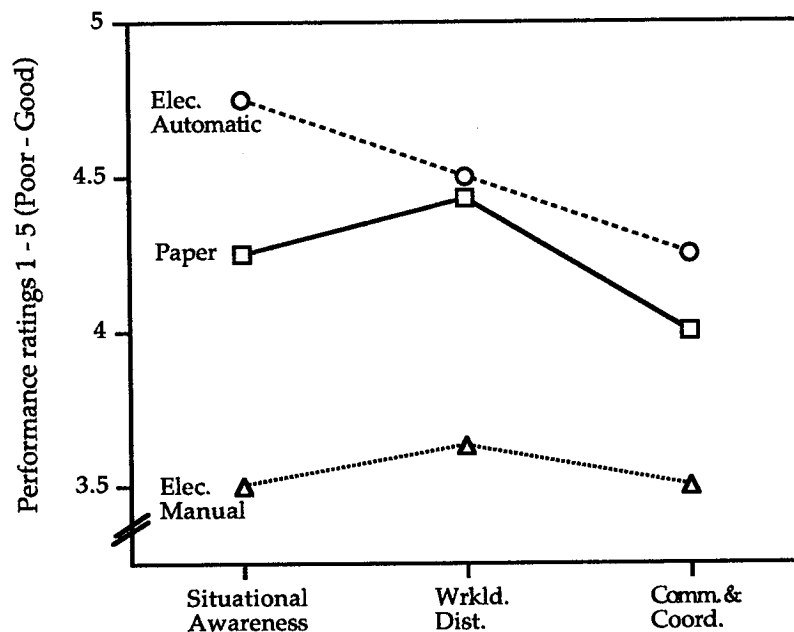


Figure 21: Performance ratings for emergency phase (individual pilots)

- *Video raters:* Following the flights, two independent raters watched the videos recorded during the sessions, rating crews and individual pilots according to a performance rating scale defined by Palmer and Degani (in preparation). Subsequent analysis of these ratings yielded no systematic differences between their respective ratings of the three groups, nor any meaningful correlation between the scores given to the same crews by the two raters. Both raters tended to give crews and individual pilots the score of "3" on a scale of 1-5, and therefore the variance in ratings was too small to enable analysis. The scale apparently did

not have enough gradations to tease-out whatever differences in performance were indeed perceived. Because of this low variance, neither rater's scores correlated significantly with the in-flight observer's performance ratings. This data were quite useless, and was not used for any further analysis.

- *Kathy Mosier*: In her paper describing the decision processes within the crews, Mosier (1992) found that the mean number of informational items discussed by the crews decreased as the checklist became more automated. This finding is interesting, as the data from the current analysis indicate that the overall number of in-cockpit communications did not differ. Since the current analysis did not look at any form of content analysis of the verbal transcriptions, and since Mosier focused her analysis on a count of speech acts with "informational value," it seems important to note the relationship between total amount of verbal communication and the "informational value" of that communication.

Mosier also found that as the automation of the checklist increased, the crews were more likely to perform the wrong action in response to the birdsrike emergency. She used the decision to shut-down an engine in response to the engine fire warning as a criteria for task performance: leaving both engines on was considered good performance, while shutting one engine was considered poor. Accordingly, she rated the Paper crews as high performers (1 of 4 crews shut down an engine), the Elec. Manual crews as medium (2 of 4 crews such down an engine), and the Elec. Automatic crews as poor (3 of 4 crews shut down and engine, though one crew immediately re-started that engine). These findings, however, seem to conflict with the current analysis of crew performance, as rated by the in-flight observer; they were not confirmed, nor opposed, by the performance ratings provided by the video raters. A comparison of crews that shut down an engine with crews that did not yielded no significant difference along non-verbal or Speech Act measures ($p > .1$). When comparing the proportion of total speech activity performed in the emergency phase of flight, the crews that shut down an engine were observed to perform a larger portion of their speech (.577) than the crews that kept both engines on (.446); this difference was statistically significant ($F_{1,11} = 29.14$; $p < .001$). Whether the increase in total speech was the cause, or result of, the inappropriate engine shutdown cannot easily be determined.

Overall, with all ratings in mind, it seems that the performance measures were not designed with sufficient care to detect meaningful differences between the three groups. The lack of variance in performance ratings by the video raters is disappointing. Further, the absence of additional ratings leaves the in-flight observer's performance ratings without support, thus

lacking a sense of perspective. Within these constraints, several interesting issues did emerge. The significantly higher ratings for the Auto is interesting, particularly since this group demonstrated an intermediate level of NVs, i.e., more than the Paper, but less than the Manual. Further, since the groups did not vary in the amount of in-cockpit speech, the strong negative relationship between performance measures and SAs suggests that the rater based his appraisal of performance on some other elements. The differences with Mosier's findings are not too surprising: her binary performance measure (engine on vs. engine off) did not really allow for much variance, and her focus on analysis of speech content did not include the various activity variables used in this study. These performance issues, along with the results pertaining to non-verbal activity and verbal communication, are discussed in detail in the following section.

5. Discussion

5.1. Non-verbal activity

The measurement and analysis of non-verbal behavior in the cockpit was the focal point of this research. The coding scheme - which examined the four visible actions of look, touch, point-in and point-out - succeeded in capturing differences between the three conditions defined by checklist design. The scheme also captured differences resulting from the different roles pilots played during the flight, i.e., the difference between PF and PNF. The fundamental difference in the three groups' overall level of activity (Figure 9) should not be surprising. The data are consistent with the connection between technology and task performance previously made (e.g., Hulin & Roznowski, 1985; McGrath & Hollingshead, 1993; Kiesler & Sproull, 1992). In an earlier section (1.3.4.), and in earlier theoretical discussions (Segal, 1990), the concept of design as choreography was proposed, suggesting that variations in interface design could directly shape an operator's physical behavior. The non-verbal data presented in Figure 9 - which was, by definition, based on physical behavior only - supports this concept. The finding of higher levels of activity in the Electronic cockpits is consistent with, and may partially explain, the findings of Wiener et al. (1991), who argue that in spite of some of the obvious advantages of the glass cockpit, the workload in a modern cockpit (in their case, an MD-88) could become excessive.

The data show that, in spite of the fact that the specific number of action items for any given checklist procedure was identical for all three groups, the two Electronic groups were observed to interact more with the system. This pattern is largely due to the Electronic crews' need to interact with the display to perform actions related to control of the checklist itself, such as call up a checklist page, acknowledge that an item was performed or clear a completed screen; pilots in the Paper group simply read items off a printed page. Recall that approximately 70% of the items on any given checklist were automated in the Auto condition; it is interesting, therefore, that the Manual group did not exhibit a significantly greater rate of activity than the Auto group. A difference in activity level and resultant use of non-verbal communication was predicted in Hypothesis 1. (section 3.3); this predicted difference was not found. The lack of difference between these two groups suggests that most of the pilots' interactions with the touch-sensitive displays concerned "meta" control of the checklist itself - e.g., paging through modes, going through menus - rather than inputs relating to checklist content.

Wiener has repeatedly argued (e.g., 1988) that automated equipment does not appear to live up to its expectations in reducing workload, since while the manual tasks may be declining, monitoring and mental workload have increased. The data presented here suggest that even at the physical level of manual interaction with the system, workload is not alleviated by automation (Figure 9). This is not surprising, as the interaction of the pilots with this particular automated system was designed to require manual interactions with the touch-sensitive displays. Both this system and the systems described by Wiener rely on manual interface for operator control of automated system, requiring manual inputs for such activities as activation, mode selection and message confirmation. Later in this section, the potential benefits and costs of the Electronic groups' added engagement with the system will be discussed.

Within each crew, the difference in the two pilots' respective roles was also clearly reflected in the non-verbal data. Looking at the two non-verbal behaviors that were directly linked to control of the checklist display - touching the display and pointing at it - the differential effect of checklist design on each pilot are apparent (Figure 10). The clear distinction between Pilot Not Flying and Pilot Flying created a situation in which one pilot - the PNF - did the majority of system control activity, while the other - the PF - was seen to be doing little other than fly the aircraft. Note that the difference in checklist design is reflected only in the PNFs' control activities, as they were the ones in charge of checklist performance. In this sense, the PFs may serve as a baseline for comparison: since the task of flying the aircraft, and the design of the flight controls, were identical for all three groups, there was essentially no difference observed in the rate of the PFs' activities in the cockpit. The PNFs, however, through interacting with either a paper checklist, or one of two touch-sensitive interfaces, displayed levels of activity which accurately reflected the design and functionality of the interface. These variation in PNFs activities seem to be the primary reason for the overall difference in cockpit activity shown in Figure 9.

Turning to the non-verbal measure that focused on the pilots' monitoring behavior - i.e., the measure of looking across the cockpit at a crewmember's checklist display - an opposite, yet somewhat similar, pattern emerges (Figure 11). Here, the difference between the groups was reflected by the significant difference in the PF's looking behavior; the PNF's did not vary in the rate of looking over at the PF's display. The data begin to suggest the role played by non-verbal, consequential, communication. Just as an earlier study of video-mediated communication identifies the value of mutual visual access in providing individuals with valuable information regarding task coordination (Heath and Luff, 1992), the "look" data

may be driven by the pilots' attempts to pick up task-relevant information. These findings are also consistent with Chapanis et al.'s observation that when multiple operators work in the same room, the modality of their communication changes, to include more visual exchange (1972); similar observations were made by Rochlin et al. (1987). Barber (1983) explains this phenomena as one operator's expectancy of another's technical competence, where monitoring is performed to confirm this expectancy. In the context of the above discussion of differences in the PNFs' activities shown in Figure 10, these visible differences may be providing the pilots from the three groups with a differential incentive to look across the cockpit at their crewmember's display: the pattern of PFs looking over at the PNFs display (seen in Figure 11) is similar to the pattern of activity displayed by the PNFs (seen in Figure 10). It should be noted, however, that due to the unclear nature of the performance data, a strong causal connection between consequential communication and quality of task performance cannot be made.

While PFs in the Manual crews spent more time looking across the cockpit than PNFs, this difference was reversed for the two other conditions. Recall that in Figure 10, PNFs in the Paper and the Auto conditions were busier manipulating the system than their respective PFs; these same pilots also did more looking across the cockpit than their PFs (Figure 11). This finding is intriguing, and suggests that the rate at which pilots looked across the cockpit, in and of itself, did not necessarily capture the differences in checklist design which were impacting the rate of display-related, non-verbal, activity. It thus seemed essential that the analysis focus on the relationship between the two measures of control inputs and monitoring. The analysis first examined the particular temporal relationship between one's look and another's activity. The lag sequential analysis confirmed that the relationship between the two behaviors was not coincidental. When looking at the entire sequence of behaviors in the cockpit, these two particular activities were shown to be strongly related. Specifically, the lag sequential analysis showed that one pilot's reach for the display was highly likely to be followed by the other pilot's turn to look at that display (Figure 12). This finding was supported with the measure of concurrence between one pilot's touch and another's look (Figure 13): on average, across all three conditions, almost half (48%) the time that PFs looked over at PNF's display, the latter were busy interacting with that display. These data nicely reflect differences between the level of activity of the PNFs in the three groups: concurrence for Paper alone was relatively low (12.5%), while for both Manual and Auto, concurrence was significantly higher (60% and 72%, respectively). Note that the ratio of the Manual and Auto PFs' concurrent looks to the Paper PFs' concurrent looks - approximately 5 to 1 - was much higher than the ratio of Manual and Auto PNFs' activities

to Paper PNFs' activities, which was approximately 3 to 1. This suggests that the higher rate of concurrent looks in the electronic cockpits was not simply a result of the higher level of activity in the cockpit, but was triggered by the visible, physical nature of the PNFs' interactions with the electronic checklist display.

In other words, in the Electronic conditions, over half of the time that PFs spent looking over at their PNF's checklist display was during the PNF's interactions with that display. Given the low overall "look" time - pilots in the three groups looked across the cockpit at average only 2.4% of the total flight time (see Figure 11) - it seems reasonable to conclude that the concurrence of "look" and "touch" was indeed not coincidental. Further, given that PFs in the Paper and Auto spent little time looking across, it is noteworthy that the measure of looking in concurrence for Auto was so much higher than that for the Paper. Hutchins and Palen (1993) argue that the gestures of pilots in a cockpit acquire their meaning by virtue of being superimposed on the meaningful spatial layout of the panel. It stands to reason that crewmembers would indeed attempt to follow each other's interactions with different panels, as these would provide real-time information pertaining to current - and changing - system states. Since a movement of the head provides a clue for a valuable source of information (Rasmussen, 1986), these data - which measured pilots' head movements to identify direction of gaze - may be providing the proof that interactions with the system were informative for crew coordination.

The relationship between "look" and "touch" is further supported by the *lack* of effect that checklist design and automation had on the time that PNFs looked across while the PFs manipulated their own display. Although, as shown in Figure 11, PNFs did look across the cockpit, it appears they were unaffected by different checklist interface designs. Further, there was no concurrence between their looks and the PF's activities (Figure 13). Given the low rate at which PFs interacted with their own displays (Figure 10), it is not surprising that the PNFs had no incentive to look at them; this, combined with the added workload that PNFs had through their own interactions with the system, resulted in effectively no concurrence between a PF's activity and a PNF's monitoring of that activity. This asymmetric flow of consequential communication - in which one operator does more "emitting," while the other does more "detecting," of non-verbal information - was accurately predicted in Hypothesis 2 (section 3.3).

The comparison of two different flight phases provides further insight to the effects of interface and automation on non-verbal crew activities. Recall that the comparison

contrasted the early phase of performing the After Start checklist - a highly routinized task - with the later phase of coping with the birdstrike, which included fault diagnosis, decision making and emergency procedures. For all three groups, the rate of activity in the normal checklist phase was much higher than in the emergency procedure (Figure 14). Since the normal phase was much shorter than the emergency phase, and since the After Start procedural task was well-structured and familiar to the crews, this finding is not surprising.

The more interesting question regarding differences across phases concerned the distribution of activity across the flight. Of all the actions taken by each crew throughout the entire flight, what portion was taken in each phase? Did checklist design and automation have a differential impact on this distribution? The data presented in Figure 15 suggests that, indeed, different interfaces resulted in different distributions of cockpit activity. The Auto crews, who performed a smaller portion of their activities in the normal checklist phase, performed a larger portion than both other groups in the emergency phase of the flight. These data are consistent with earlier aviation field studies (Curry, 1985; Wiener, 1985; Wiener, 1989), where pilots in more "modern" cockpits experienced that while automation did help in low-workload situations, it actually increased the workload in high workload situations. Similar results were obtained by Costley et al. (1989), who compared crew activity in the conventionally instrumented B737-200 with the modernized and automated B737-300. They had found that in high-workload segments of the flight, crews in the automated planes showed higher rates of activity (151% as much), while in the lower workload phase they were slightly less active (90%).

Reviewing the non-verbal activity data, the overall pattern of greater observed activity in the Electronic cockpits is quite clear. So is the effect of pilot's role on the particular activities of display manipulation and looking across the cockpit, as well as the strong connection between display manipulation by one pilot and their being monitored by the other pilot. These behaviors may impact crew performance in two different ways. On the one hand, greater visual access of pilots to their crewmembers' activities may promote better cooperation. This type of crew dynamic has been suggested by McLeod (1992), who compiled six studies of the effects of Group Support System on group process, finding that the weight of the evidence suggested that GSS increases equality of participation and decreases domination by a few individuals. Similarly, the Electronic checklist design, by causing more interaction between the PNF and the interface, provided the PF with the opportunity to monitor what was being done and become more involved in that process. This may be reflected by the relative increase in speech in the decision-making emergency

phase, a phenomena discussed further below. From this perspective, the choreography of greater cockpit activity in the Electronic conditions may be seen as supporting information exchange in the cockpit, with potential enhancement of crew cooperation.

On the other hand, one must acknowledge the possible negative effects that highly visible activity may have on crew performance. Is it possible that the PNF's increased control activity distracts the PF's attention away from their primary task of flying the aircraft? The more group members share all of the information they have, the more likely it is that they will suffer serious information overload (McGrath and Hollingshead, 1993). Recall the ASRS example #7 described earlier (section 1.3.4.), in which the FO was so busy watching the Captain's actions that he lost track of his own task. From this perspective, the choreography of operator activities should take into consideration not only the importance of rendering certain activities visible, but also the importance of *not* making activities that are not important too distracting.

5.2. Speech Acts

Wiener (1993) describes an observation he made in an advanced cockpit (757) on a flight from Atlanta to Washington National, in which from climb-out until preparation for arrival, not a word was spoken; the only intra-cockpit communication he could observe was between each pilot and his CDU. How does the current data support his observation?

The most prominent finding regarding the rate of intra-crew speech acts was the *lack* of any difference between the three experimental conditions (Figure 16). Whether looking at rate or duration of speech acts, no systematic difference was found, even though the coding of non-verbal activities in the cockpit did show that the three checklist designs elicited significantly different levels of non-verbal in-cockpit activity (Figure 9). Given this difference in activity, it seemed reasonable to expect that crews in the two Electronic conditions - who displayed significantly more activity than Paper crews - would engage in more verbal exchange; a concurrent increase in activity and speech was shown in previous studies (e.g., Strauss and Cooper 1989; Costley et al., 1989). Alternatively, a strong argument for the benefits of non-verbal communication might suggest that crews who are more active should exhibit *less* speech activity. This type of relationship between activity and speech is described by Linde and Shively (1988), who found that crews modulated the frequency of intra-crew communication according to the workload imposed by the task scenario: when workload was high, speech was cut down to a minimum, while low

workload enabled the pilots to fill the auditory space. A similar argument was made by Bowers et al. (1993), whose interpretation of communication and performance data suggests that crews might utilize workload savings to communicate more frequently.

The finding of no difference in overall speech rate suggests an "attractor state" in the wide range of speech-act frequencies that may be exhibited during such a flight. In this respect, the particular flight scenario may be constraining the crews to perform similar rates of speech acts, regardless of the non-verbal activity in the cockpit. The relationship between speech and activity, captured by the measure of NV/SA (Figure 18), essentially reflects the difference in activity described in Figure 9, as well as the effect of role described in Figure 10. It thus seems that the environment - including the difference in checklist interface and automation - did not directly impact the verbal component of crew interactions. In other words, the verbal communication component of the global task remained similar in spite of variations in the demands of non-verbal components of the task. Wiener (1993) does suggest that since the demands of the modern cockpits may simply be qualitatively different, higher amounts of conversation may be neither necessary nor desirable. If his use of "higher" is in any way related to the fact that automation does seem to cause a higher level of activity, the current data seem to support his assertion.

The distribution of speech across the two different flight phases of normal and emergency procedures (Figure 17) - and the particular interaction effect of phase and condition - reflect a pattern discussed earlier regarding distribution of non-verbal activity. As was evident in the analysis of non-verbal activity across flight phases, we see that the Auto crews performed the smallest proportion of speech in the standard procedure phase, and the largest, in the emergency phase. Results from research in Group Support Systems (GSS) may provide a basis for interpretation of these data. There is considerable evidence that it is much more difficult for groups to reach consensus using electronic communication, compared to face-to-face groups (McGrath & Hollingshead, 1993). This argument was further supported by McLeod (1992) in a review of four lab studies which included 96 groups, in which all findings of measured time to consensus, or time to decision, indicated that technology-supported groups took longer than unsupported groups. Similarly, in their study of crews in conventional and automated cockpits, Costley et al. (1989) found that in high workload conditions, frequency of crew communication was higher in the automated 737-300 cockpit by as much as 27%. What we see in the current data is a relatively larger increase in speech rates in the automated cockpit when the task required the resolution of a problem and management of an emergency procedure. Could the pilots be discussing

control of the automation itself, rather than the actual problem created by engine fire? The lack of difference in non-verbal activity between the Manual and Auto crews suggests that much of the activity was indeed directed at management of the automation, rather than at performing particular items in a procedure. Although a thorough analysis of speech content would be necessary to determine what specifically was discussed by these crews in each phase, the relative difference in distribution of speech acts across normal-procedure and emergency phases suggests that automation - along with the context provided by the problem solving scenario - did indeed impact the organization of crews' verbal performance, as it did their non-verbal activity.

From a methodological perspective, the lack of overall difference between the rate of speech displayed by the three groups supports the theoretical argument for inclusion of non-verbal behavior in the measurement and study of crew communication. In spite of the fact that the different groups differed significantly in their level of activity - and, to a certain extent, in their task performance - these different levels of activity and performance were not accompanied by an increase nor a decrease in rate or duration of intra-crew speech communication. At the same time, it seems clear that pilots did direct their attention at crew-member activities, suggesting that non-verbal information was flowing in the crew. The data thus suggest that previous studies that looked at verbal communication alone (e.g., Kanki et al., 1989; Foushee & Manos, 1981) may have benefited from including cockpit activity in their observations.

5.3. Performance

The process of measurement and analysis of crew performance was greatly compromised by the "piggy-back" nature of this study. Since the current analysis is based on video data collected by Palmer and Degani (1991), and since the rating of performance by expert observers was carried out prior to the proposal of this study, the author had little control over the design of performance scales and actual ratings by the observers. Nevertheless, there seem to be sufficient sources of performance evaluation from which to draw initial conclusions.

In an elaborate experiment comparing crew performance under different levels of cockpit automation and carried out by Wiener et al. (1991), the three measures of crew performance - as judged by expert raters, instructors and observers - produced few statistically significant contrasts. The data presented here seem to follow the same pattern: of the three expert

performance raters, only the in-flight observer's ratings seemed to capture a systematic difference between the three experimental groups. The rating data for performance over the entire leg, presented in Figure 20, suggest that the Electronic interface promoted better crew communication, management style and coordination. Of the two Electronic groups, automation seemed to promote higher performance ratings. The ratings focusing on performance during the emergency phase (Figure 21) seemed to reverse the position of Manual and Paper crews, while continuing to show a slight advantage for the Auto crews. These trends conflict with Wiener et al.'s findings (1991), where, whenever significant performance results were found, they favored the less automated DC9 crews over the automated MD88 crews. Other studies have also found that under high task difficulty, pilots in automated conditions were worse at problem solving than in the manual condition (Thornton et al., 1992; Bowers et al., 1993; Mosier, 1992).

Since Mosier's data was taken from the same video tapes, her findings are most interesting. Recall that her definition of performance was based on whether the crew shut down an engine following the birdstrike; a shut down was considered "wrong." Of the twelve crews, six did shut down; five of these six were flying the Electronic interface. The erroneous tendency of the more "hi-tech" Electronic crews to shut down an engine is consistent with the GSS data presented by Kiesler and Sproull (1992), who found that groups that met through computers were somewhat risk seeking in all circumstances of decision making. Since the simulated aircraft had only two engines, the decision to shut down one of them can certainly be seen as risky. How, then, can the conflict between these data and the in-flight observer's ratings be resolved?

Obviously, the observer's data may simply be wrong. This seems to be an easy solution, albeit one that does not account for the observer's high level of expertise and extensive experience as a performance evaluator for NASA. There may be an alternative way of interpreting the data, though. Dividing the 12 crews to two groups along Mosier's performance measure regarding engine shut down (1992) showed that the low-performing crews exhibited a significantly larger proportion of speech acts in the emergency phase. This seems consistent with the finding of negative correlations between performance measures and rate of speech acts, presented in Table 2. While the lack of difference between the three groups' *global* speech rate suggests that the relationship between performance and speech may be defined by other variables that were not included in the measurement and analysis process, the distribution of speech across different flight phases suggests that the

in-flight observer may have been picking up *local* differences in speech and performance which were consistent with Mosier's findings.

It is essential to note, though, that performance rating according to one criteria only, such as the "engine shut down" measure used by Mosier, clearly does not reflect the complexities involved in performance of this type of task. For example, a strong argument can be made that since the birdstrike occurred immediately after takeoff, a quick return for landing is the most obvious correct response to the problem. In fact, many pilot training programs emphasize the advantages of, when possible, solving problems on the ground rather than in mid-air. From this perspective, the crews who returned *fastest* for landing performed best. If one follows this assumption, rank-ordering the 12 crews according to total legtime produces an entirely different performance picture: of the six fastest crews to return to landing, five were from the Electronic interface conditions (3 from Manual and 2 from Auto). This illustrates the lack of stability of *any* binary performance measure and, once again, suggests that more thorough analysis of performance is needed before conclusions are drawn.

5.4. Team Engagement State Space (TESS) - a qualitative interpretation

Figure 22 presents a simple application of the TESS model proposed earlier (section 1.3.3.) to the analysis of the non-verbal and verbal video data. Recall that, in the initial description of TESS, the intention was to describe a crew's performance in a way that captures the interaction between the control and the communication tasks. Since the promotion of non-verbal activity as communication argues that the control and communication are integrated, the variables plotted on each axis reflect the assumptions that integration was indeed captured by the data. The vertical axis (Y) shows the rate of observed non-verbal cockpit activity, i.e., the direct interactions of individual pilots with the display. Based on the assumption that certain types of non-verbal activity are specifically informative, the horizontal (X) axis shows the sum of what were assumed to be "communicatory" behaviors: rate of speech, rate of pointing and rate of looking over at a crewmember's activity. Each data point reflects these two measures for each pilot, whether PF or PNF. The data points are coded by shape to reflect the particular condition, or checklist design, to which that pilot belonged; a "•" symbol identifies the data points for PFs. The legend provides the mapping between symbol and group; note that group means are also displayed.

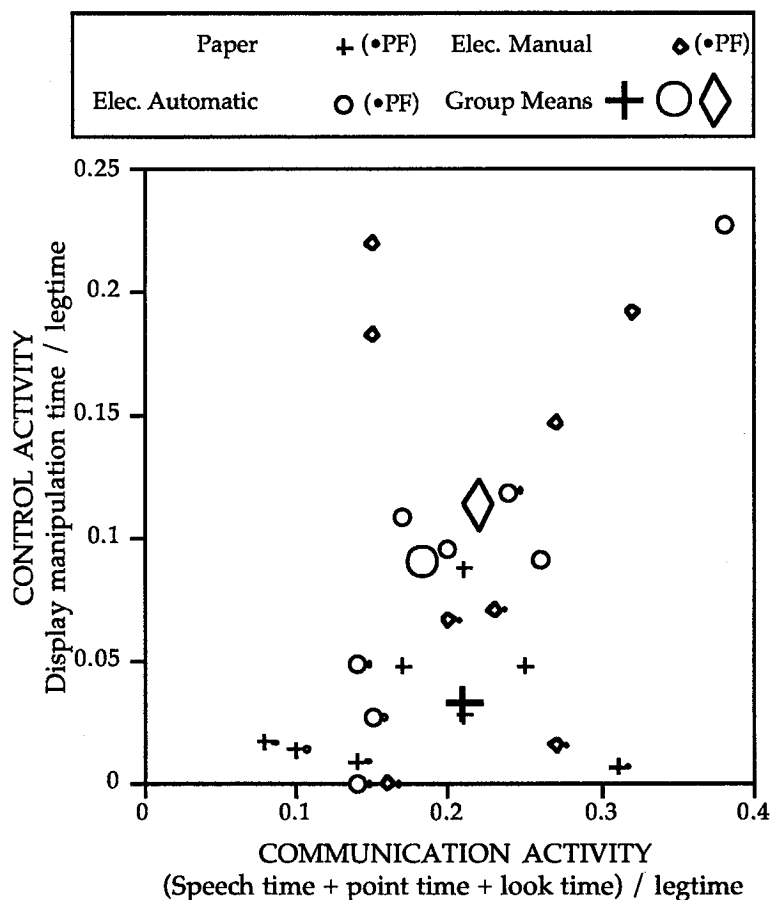


Figure 22: Application of TESS model for data analysis

Looking first at the group means, the plot reflects one of the most interesting findings concerning the effects of interface design on task performance and crew coordination. The Electronic interface created a task environment in which pilots performed more interactions with the system without a noticeable increase or decrease in communication activity. Both of these groups show greater variance in control activity than the Paper, while the variance in communication activity for all three groups is very similar. The Paper pilots seem to differ primarily along the communication component of the task; it seems that the paper checklist did not cause substantial differences between pilots in the system control task.

Figure 23 provides an interesting perspective on the distribution of activity between the two crew members, by connecting the data-point for PF and PNF for each crew. Note that for the Manual crews, the unequal distribution of control activity between the two pilots was quite clear, as can be seen from the strong vertical component of the vectors connecting the

crew members. At the same time, the Manual pilots show little variance in communication activity. The Paper crews - as a result of less variance in control activity, and more variance in communication activity - have the shallowest vectors; the Auto crews, an intermediate pattern. From this perspective, it seems that the design of the conventional Paper checklist and the Automated electronic checklist resulted in a more equal distribution of control activity than did the hybrid electronic Manual checklist. Note also that for most crews, the slope is positive, i.e., the PNFs were doing more of both control and communication activities.

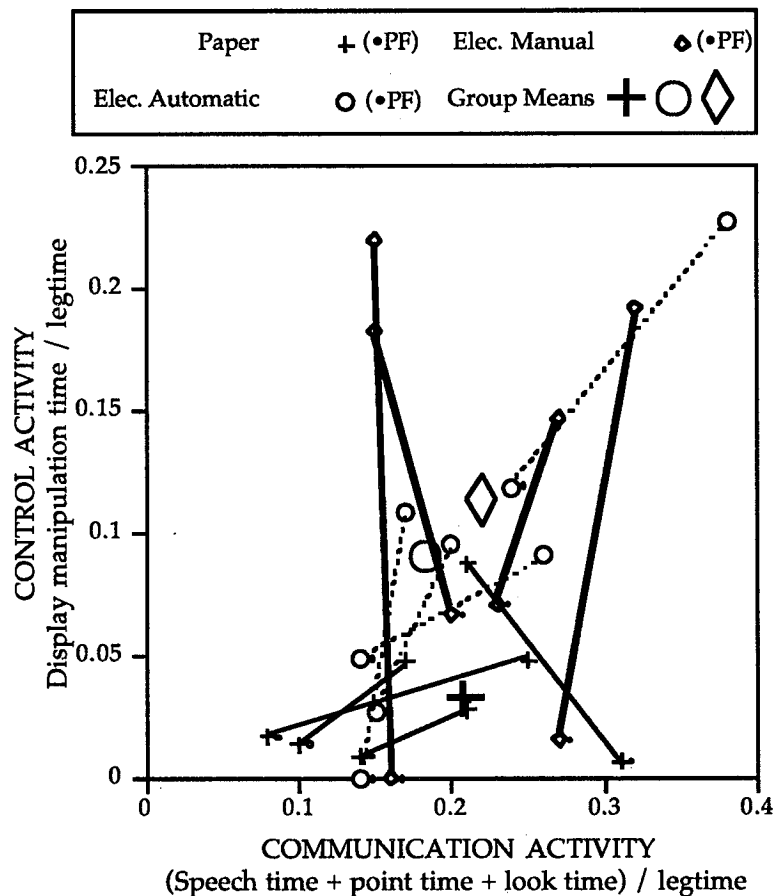


Figure 23: Relationship between PF and PNF in the TESS graphic

This cursory analysis highlights the advantages and drawbacks of modern interface designs and automation. The conventional paper checklist seems to allow the task to be performed with minimal interactions with the system, thus reducing the opportunity for input error. Indeed, the shallow nature of the Paper crews' vectors seems to be partially a result of an apparent "floor effect" in control activity. Automation of the checklist increases input

activity by virtue of imposing on the pilots the task of interacting with the interface in the process of "communicating" with the automated system; at the same time, it provides an added level of security, since the system continuously tracks the pilots' performance of the checklist procedures. Simply providing a modern interface without automation - which was the case with the Manual - increases the workload of the control task to the highest degree, but does not afford the advantages inherent in automation.

Recall, though, that the added activity exhibited by the two electronic groups seemed to trigger cross-cockpit looks from crew members. This suggests that while workload may indeed have been higher in the Manual condition than in the Paper condition, the Manual crews may have benefited from the redundancy afforded by the PFs' monitoring of the PNFs' interactions with the checklist display. Since performance measures did not identify errors of input - and recovery from such errors - this assumption needs to be further tested. Overall, a visual analysis of the TESS suggests that variations in interface design and automation seem to impose increased workload on manual activity without an equal increase in the load on the crews to communicate and coordinate task performance. Insofar as increased manual workload may afford information that can be observed by a crew member, it may enhance the overall reliability of crew performance.

5.5. Potential applications and future research

The results presented above seem to support the two primary thrusts of the theoretical concepts which inspired this research. Differences in checklist interface design resulted in a clear difference in non-verbal activity between the Paper and the two electronic conditions. This difference validates the notion of designer as choreographer. This, in and of itself, would not be too important had the results not shown that pilots do indeed monitor each other's non-verbal interactions with the system. The strong connection between one pilot's activities and the other's monitoring suggests that the choreographed control "dance" does indeed contain information that is noted by the pilots. Thus, the findings support the notion that it is incumbent upon designers to consider the impact that interface design might have on crew coordination.

Since the activity of one operator may cause another to look over at the location of activity, several broad design guidelines may be defined. It would certainly be useful to build upon this phenomena in order to encourage the monitoring of actions that may have critical consequences on the task. Hence, if a task analysis identifies certain actions that demand

not only the attention of one operator, but rather, the redundant attention of other operators, such actions should be choreographed to take place in a manner and location that afford visual access to the entire crew. Conversely, the attentional demands of monitoring may be distracting to operators involved in the performance of critical task components. From this perspective, it seems important to identify those task elements that are so uncritical that they need not be monitored, and to design them so that they do not impose added workload on those operators who are not directly in charge of their completion.

The "informational value" of observed non-verbal activity needs to be identified and defined. This could possibly be done by constructing scenarios where particular actions by one operator are designed to elicit a specified response from another operator; thorough debriefing and interviews may provide further information regarding the semantic value afforded by observing different activities. It also remains to be determined at what accuracy, or resolution, pilots actually pick up control manipulations. One of the drawbacks of touch-sensitive technology is that it renders all control activities similar, namely, all controls involve pressing the display. Under these conditions, switching a system from "off" to "on," or vice versa, entail identical movements on part of the operator. If the displays are not designed to explicitly make the resultant changes in the system visible and discriminable, the touch-sensitive interface will eliminate the variability in control actions which exists in traditional cockpits, e.g., flip-up vs. flip-down, pull vs. push, turn vs. slide (for a discussion of the importance of visibility in design, see Norman, 1988). To the extent that such subtle control motions are perceived and used by other pilots, the particular designs used in this experiment did not enable information at this level of detail to emerge. Measurement and analysis at this scale can be expected to involve sophisticated video technology, as well as rigorous coding procedures that may capture differences in pilot-control interactions and crewmember reactions. It seems that such a procedure would be possible only under simulated flight conditions, where both system design and task scenario provide the researchers with maximum, and precise, control.

The findings support the argument that designing multi-operator crew-machine interfaces is fundamentally different from the design of single-operator human-machine interfaces. As the majority of applied aviation human factors research and data focus on the measurement and analysis of the interaction between a single pilot and his/her cockpit interface, it seems essential that the unique aspects of designing for crews be studied and established in greater depth. Since the study of cockpit crew factors, e.g., crew dynamics and decision making processes, has focused primarily on the measurement and analysis verbal interactions,

resultant findings have focused primarily on training, rather than on concrete design issues. The inclusion of non-verbal measures as fundamental elements in the process of crew communication will enable designers of large, multi-operator, cockpits to directly benefit from crew factors research.

Obviously, the current analysis was highly constrained by the fact that the experiment was not designed to explore non-verbal communication. One clear drawback was the location of the Manual and Auto groups' checklist interface, at the bottom half of the touch-sensitive CRT, underneath the system-control interface that was used by all three groups. Although the Paper crews held the checklist in their hand, and had no electronic interface with the checklist itself, their interaction with the top part of the display was reflected in the non-verbal activity data. Any experiment that sets out to focus on the different impact of paper and electronic interfaces should be designed to minimize confounding the two, by making those aspects of the task that are being compared spatially unambiguous; the checklist interface should have been presented on a dedicated, exclusive, touch-sensitive display. Such design of mutually exclusive interfaces would also afford precise design of measurement procedures, in particular, optimal camera location and precise operational definition of different control actions.

The method of measurement of non-verbal activity and its potential contributions to crew communication proved to be complex. Placement of the video cameras behind the pilots in the aircraft's centerline - which provided observers with an over-the-pilots'-shoulders perspective - provided reasonable resolution regarding reaching and touching the checklist display. Other activities, however, such as manipulations of switches located on the pedestal or off to the side, were either occluded by the pilots or simply outside the cameras' field of view. It seems that two additional wide-field-of-view cameras mounted above each pilot's head may provide better coverage of the cockpit, and a perspective that is closer to that which the pilots themselves occupy. Further, the behind-the-back perspective did not provide accurate information regarding the pilots' focus of visual attention; while the coding scheme captured those times when pilots clearly turned their heads to look across the cockpit, it can be assumed that quick glances that were performed without head movement had gone undetected by the coders. To truly capture this aspect of the pilots' in-crew monitoring activity, eye-trackers would have provided invaluable data, which, combined with manual system-control activity in the same time-line, could serve to accurately discriminate the informativeness of specific actions and particular locations of activity.

The coding of non-verbal activity from the video recordings proved equally challenging. Ideally, one would want the coding to be done by observers who are as familiar with the task domain as the pilots themselves. In this respect, one would want to allocate research funds to obtain pilots not only as subjects, but also for assistance with all subsequent coding and analysis tasks. The procedure of "coding by consensus," though time consuming, seems to provide data that are more consistent and reliable than independent coding. These two requirements emphasize the logistic difficulty of obtaining a large number of highly skilled individuals for the purpose of data interpretation and coding. The key issue remains the operational definition of particular actions. It may be useful to obtain subjective reports from the pilots themselves regarding the visibility and informativeness of different cockpit activities. Hopefully, the inclusions of eye-trackers in the measurement process will further clarify those specific actions and locations that draw the pilots' attention, thus providing further validity to the coding scheme.

While certain implications of particular designs were discussed above, the findings failed to provide the necessary connection between design and performance, thus precluding specific conclusions regarding specific design recommendations. Since the performance ratings given by the in-flight observer conflict with the analysis of performance suggested by Mosier (1992), it is hard to assign concrete values to any one of the three different checklist interfaces. Further, since a clear performance advantage could not be identified, it is impossible to provide specific design guidelines which relate to checklist design. It is interesting that the literature (e.g., Wiener, 1991) describes numerous similar cases, in which the comparison of manual and automated systems failed to yield significant differences in performance ratings. The repeated failure of applied research to reveal significant effects highlights the fundamental difficulty associated with this type of work, namely, the difficulty of assembling a sample that is large enough to provide sufficient statistical power. As one gets closer to "real world" settings, the procedures of recruiting highly proficient subjects become extremely complex - and, more importantly (and unfortunately), extremely expensive. While these complications do not diminish the importance of applied research, they should encourage researchers to be creative in the use of basic research settings as springboards for applied work.

Nevertheless, discussion of the data presented in the TESS does suggest that in the context of this experiment, the Manual condition seemed to be the poorest design, since it significantly increased cockpit activity without offering any benefit associated with automation. If the Auto design were to *reduce* workload below Paper, one could argue

against it on the basis that it may cause the crews to remain "out of the loop." This, however, was not the case: overall, crews in the Auto condition were busier than the Paper crews. In this respect, changes in interface design alone, such as the change from a hand-held, paper checklist to an electronic-manual interface, do not seem justify the cost in materials and training which are usually associated with such upgrades. Further, the lack of difference in activity level between the Auto and Manual groups suggests that automation at the content level alone - i.e., automation that performs only the items on a particular checklist, while leaving the task of choosing the checklist to the pilots - does not necessarily alleviate workload. While the Auto checklist may have provided added redundancy, for automation to truly impact workload, all levels of the task - in this case, the checklist task - should be automated, including such things as mode selection and identification of appropriate page/procedure. The argument can be made that an upgrade of the functionality of the system, such as provided by automation, is a necessary element in any interface modernization program.

6. Summary

No one has a more sweeping influence on task performance - and crew-machine interaction - than the designer (Senge 1990). This statement presents the applied objective which this research proposal wished to address. Specifically, the proposed research hoped to identify some of the unique qualities of *crew-machine interaction*, i.e., how multiple operators interact in the context of performing a control task. Using video data recorded in a high-fidelity simulator, the research program provided a methodology for coding and analyzing the complex relationship between pilots involved in the task of controlling an aircraft and managing its systems under high-workload conditions. The primary goal of the research was to identify connection between control activity and crew coordination; this was done by including non-verbal communication as an integral part of information exchanged in the cockpit. Based on the analysis of video and audio recordings, the results identified several consistent differences between crews flying aircraft with different checklist interface designs. The research program aimed to unite - in both theory and application - two streams of human factors research. In this respect, the results provided the connection between studies of group dynamics and the design of human-machine interface. It is hoped that the results encourage further investigations that look at the non-verbal aspect of communication, and that they may serve to define specific guidelines for the design of future cockpits and similar multi-operator control stations.

References

(ASRS) Aviation Safety Reporting System (1991). Non-Verbal Flight Crew Coordination Incidents. Search Requests No. 2277 & 2337. ASRS Office, Mountain View, California: Battelle.

Alcock, John (1989). Animal Behavior (Fourth Edition). Sinauer Associates Ltd; MA.

Ashby, W. Ross (1956). An introduction to Cybernetics. Methuen & Co., New York, NY.

Barber, (1983). The Logic and Limits of Trust. Rutgers University Press; New Brunswick, NJ.

Bateson, G. (1972). Steps to an ecology of mind. Random House Inc.; New York, NY.

Bateson, G. (1979). Mind and Nature: A necessary unity. E.P. Dutton, New York, NY.

Birdwhistell, R.L. (1970). Kinesics and Context. University of Pennsylvania Press.

Bowers, C., Deaton, J, Oser, R., Prince, C., and Kolb, M. (1993). The Impact of Automation on Crew Communication and Performance. Proceedings of the 7th International Symposium on aviation Psychology. Ohio State University; Columbus, OH.

Burgoon, J.K. (1980). Nonverbal communication in the 1970s: An overview. In D. Nimmo (Ed.), Communication yearbook 4. Transaction; New Brunswick, NJ.

Burgoon, J.K., Buller, D.B., & Woodall, W.G. (1989). Nonverbal Communication: The Unspoken Dialogue. Harper & Row Publishers, Inc.; New York, NY.

Chapanis, A., Ochsman, R. B., Parrish, R. N., and Weeks, G. D. (1972). Studies in Interactive communication: I. The Effects of Four Communication Modes on the Behavior of Teams During Cooperative Problem-Solving. Human Factors, 14.

Chapanis, A. (1991). To Communicate the Human Factors Message, You Have To Know What the Message Is and How To Communicate It. Human Factors Society bulletin, 34/11.

Chidester, T. R., Kanki, B., and Helmreich, R. L. (1989). Performance Evaluation in Full Mission Simulation: Methodological Advances and Research Challenges. Proceedings of the Fifth International Symposium on Aviation Psychology; R. S. Jensen (Ed.). Ohio State University.

Costley, J., Johnson, D. & Lawson, D. (1989). A Comparison of Cockpit Communication B737 - B757. In R.S. Jensen (Ed.), Proceedings of the Fifth International Symposium on Aviation Psychology. Ohio State University.

Ekman, P., & Friesen, W. V. (1976). Measuring Facial Movement. Environmental Psychology and Nonverbal Behavior 1(1).

Emlen, S.T. (1975). The stellar-orientation system of a migratory bird. Scientific American, 223.

Foushee, H. C. (1982). The Role of Communication, Socio-Psychological, and Personality Factors in the Maintenance of Crew Coordination. Aviation, Space, and Environmental Medicine, 53(11).

Foushee, H. C., & Helmreich, R. L. (1988). Group Interaction and Flight Crew Performance. In E.L Wiener & D.C. Nagel (Eds.) Group Interaction and Flight Crew Performance. San Diego: Academic Press Inc.

Foushee, H. C., & Manos, D. (1981). Within cockpit communication patterns and flight performance. In C.E. Billings & E. Cheany (Eds.), Information Transfer Problems in the Aviation System. NASA TP 1875.

Foushee, H. C. (1984). Dyads and Triads at 35,000 Feet. American Psychologist, Vol. 39, 8.

Friesen, W. V., Ekman, P., & Wallbott, H. (1979). Measuring Hand Movements. Journal of Nonverbal Behavior 4(2).

Gibbs, R., & Muller, R.A.G. (1990). Conversation as coordinated, cooperative interaction. In S.P. Robertson, W. Zachary, & J.B. Black (Eds.), Cognition computing, and cooperation. Ablex; Norwood, NJ.

Goffman, E. (1981). Forms of Talk. University of Philadelphia Press; Philadelphia, PA.

Gottman, J.M., & Roy, A.K. (1990). Sequential Analysis: A Guide for Behavioral Researchers. Cambridge University Press; New York, NY.

Haslett, J.B. (1987). Communication: Strategic Action in Context. Lawrence Earlbaum Associates; Hillsdale, NJ.

Hawkins, F. H. (1987). Human Factors in Flight. Gower Technical Press; UK.

Heath, C., and Luff, P. (1992). Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction. Human Computer Interaction, 7. Lawrence Earlbaum Associates, Inc.

Helmreich, R.L. (1982). Pilot Selection and Training. Paper presented at the meeting of the American Psychological Association, Washington, D.C.

Hulin, C.L., & Roznowski, M. (1985). Organizational Technologies: Effects on Organizations' Characteristics and Individuals' Responses. Research in Organizational Behavior, 7.

Hutchins, E. (1989). The technology of team navigation. In J. Galegher, R. Kraut, & C. Edigo (Eds.), Intellectual Teamwork: Social and Technical Bases of Cooperative Work. Lawrence Earlbaum Associates; Hillsdale, NJ.

Hutchins, E., and Palen, L. (1993). Constructing meanings from space, gesture, and talk. Draft paper prepared for the NATO sponsored workshop on "Discourse, tools and reasoning: situated cognition and technologically supported environments." November 2-7, Lucca, Italy.

Ivergard, T. (1989). Handbook of Control Room Design and Ergonomics. Taylor and Francis Ltd.; UK.

Kanki, B. G., Greaud, V. A., & Irwin, C. M. (1989a). Communication Variations and Aircrew Performance. In R.S. Jensen (Ed.), Proceedings of the Fifth International Symposium on Aviation Psychology. Ohio State University.

- Kanki, B. G., Lozito, S., & Foushee, C. H. (1989b). Communication Indices of Crew Coordination. Aviation, Space, and Environmental Medicine, 60, 1.
- Keeton, W.T. (1974). The mystery of pigeon homing. Scientific American, 231.
- Kennedy, J.J. (1983). Analyzing Qualitative Data: Introductory Log-Linear Analysis for Behavioral Research. Praeger Publishers; NY.
- Kiepenheuer, J. (1985). Can pigeons be fooled about the actual release site position by presenting them information from another site? Behavioral Ecology and Sociobiology, 18.
- Kiesler, S., and Sproull, L. (1992). Group Decision Making and Communication Technology. Organizational Behavior and Human Decision Processes, 52. Academic Press, Inc.
- Klein, G.A. (1989). Recognition-primed decisions. Advances in Man-Machine Systems Research, 5.
- Kreckel, M. (1981). Communicative acts and shared knowledge in natural discourse. Academic Press; New York, NY.
- Larson, C.E., & LaFasto, F.M.J. (1989). Teamwork: What Must Go Right / What Can Go Wrong. Sage; Newbury Park, CA.
- Lave, J. (1988). Cognition In Practice. Cambridge University Press; UK.
- Levinson, S.C. (1983). Pragmatics. Cambridge University Press; UK.
- Linde, C., & Shively, R. J. (1988). Field Study of Communication and Workload in Police Helicopters: Implications for AI Cockpit Design. Proceedings of the Human Factors Society 32nd Annual Meeting. Santa Monica, CA.
- Malinowski, B. (1923). The problem of meaning in primitive languages. In C.K. Ogden & I.A. Richards (Eds.), The Meaning of Meaning. Routledge and Kegan Paul; London.

McGuire, T.W., Kiesler, S., and Siegel, J. (1987). Group and Computer-Mediated Discussions Effects in Risk Decision Making. Journal of Personality and Social Psychology, 52. American Psychological Association, Inc.

McGrath, J.E. (1964). Social Psychology: A brief introduction. Hole, Rinehart & Winston; New York.

McGrath, J.E., & Hollingshead, A.B. (1993). Putting the "Group" Back in Group Support Systems: Some Theoretical Issues About Dynamic Processes in Groups with Technological Enhancements. In L.M. Jessup & J.E. Valacich (Eds.) Group Support Systems: New Perspectives. N.Y.: McMillan.

McLeod, P.L. (1992). An Assessment of the Experimental Literature on Electronic Support of Group Work: Results of a Meta-Analysis. Human-Computer Interaction, 7.

Miller, G. A. (1973). Communication, Language and Meaning. New York: Basic Books.

Mosier, K.L., Palmer, E.A., & Degani, A. (1992). Electronic Checklists: Implications for Decision Making. Proceedings of the 36th Annual Meeting of the Human Factors Society. HFS; Santa Monica.

Nagel, D.C. (1988). Human Error in Aviation Operations. In E.L Wiener & D.C. Nagel (Eds.), Human Factors in Aviation. San Diego: Academic Press Inc.

Nickerson, R.S. (1981). Some characteristics of conversations. In B. Shackel (Ed.), Man-Computer interaction: Human factors aspects of computers and people. Sijthoff and Noordhoff; Alphen aan den Rijn, The Netherlands.

Norman, D.A. (1988). The Psychology of Everyday Things. Basic Books, Inc.; New York.

Olsen, S.E., & Rasmussen, J. (1989). The Reflective Expert and the Prenovice: Notes on Skill, Rule- and Knowledge-based Performance in the Setting of Instruction and Training. In L. Bainbridge & S. A. R. Quintanilla (Eds.), Developing Skills with Information Technology. John Wiley & Sons, Ltd.

Olson & Olson (1992). Introduction to This Special Issue on Computer-Supported Cooperative Work. Human-Computer Interaction, 7.

Palmer, E.A., & Degani, A. (1991). Electronic Checklists: Evaluation of two levels of automation. In R. Jensen (Ed.) Proceedings of the Sixth Symposium on aviation Psychology. Ohio State University; Columbus, OH.

Palmer, E.A., & Degani, A. (in preparation). Electronic Checklists: A system evaluation in a full mission simulation. NASA-Ames Research Center Technical Report; Moffett Field, CA.

Papanek, V. (1985). Design for the Real World: Human Ecology and Social Change. Academy Chicago Publishers; Chicago, IL.

Perrow, C. (1983). The Organization Context of Human Factors Engineering. Administrative Science Quarterly, 28.

Pollack, I. & Madans, A. B. (1964). On the Performance of a Combination of Detectors. Human Factors, October.

Rasmussen, J. (1986). Information Processing and Human Machine Interaction. Elsevier Publishing Co., Inc., New York, NY.

Reason, J.T. (1990). Human Error. Cambridge University Press.

Rochlin, G.I., LaPorte, T.R., & Roberts K.H. (1987). The self designing, high reliability organization: Aircraft carrier flight operations at sea. Naval War College Review.

Sanderson, P.M. (1991). Exploratory Sequential Data Analysis (ESDA). EPRL Technical Report; University of Illinois at Urbana-Champaign.

Sanderson, P.M. (1993). Designing for simplicity of inference in observational studies of process control: ESDA and MacSHAPA. Proceedings of the Fourth European Conference on Cognitive Science Approaches to Process Control (CSPAC'93): Designing for Simplicity. August 25-27, Copenhagen, Denmark.

- Segal, L. D. (1989). Differences In Cockpit Communication. In R.S. Jensen (Ed.), Proceedings of the Fifth International Symposium on Aviation Psychology. Ohio State University.
- Segal, L.D. (1990). Effects of Aircraft Cockpit Design on Crew Communication. In E.J. Lovesey (Ed.), Contemporary Ergonomics 1990. Taylor & Francis; UK.
- Segal, L.D. (1993). Automation Design and Crew Coordination. Proceedings of the 7th International Symposium on aviation Psychology. Ohio State University; Columbus, OH.
- Senge, P.M. (1990). The Fifth Discipline: The Art and Practice of the Learning Organization. Doubleday/Currency; New York, NY.
- Sexton, G. A. (1988). Cockpit-Crew systems Design and Integration. In E.L Wiener & D.C. Nagel (Eds.), Human Factors in Aviation. San Diego: Academic Press Inc.
- Shaffer, M. T., Hendy, K. C. & White, L. R. (1988). Empirically Validated Task Analysis (EVTA) of Low Level Army Helicopter Operations. Proceedings of the Human Factors Society 32nd Annual Meeting, Vol. 2.
- Small, R.L., Flory, D.E., Munger, M.P., Williamson, D.T., & Hollis, B.T. (1988). A Cockpit Natural Language Study: Selected Transcripts. Midwest Systems Research Inc.; Dayton, OH.
- Smith, M. & Wilson, E. A. (1963). A Model of Auditory Threshold and its Application to the Problem of Multiple Observers. Psychol. Monog., 359.
- Stone, R. B. & Babcock, G. L. (1988). Airline Pilots' Perspective. In E.L Wiener & D.C. Nagel (Eds.), Human Factors in Aviation. San Diego: Academic Press Inc.
- Strauss, S. G., & Cooper, R. S. (1989). Crew Structure, Automation and Communication: Interaction of Social and Technological Factors on Complex Systems Performance. Proceedings of the Human Factors Society 33rd Annual Meeting.

Thornton, C., Braun, C., Bowers, C., and Morgan, B.B. (1992). Automation Effects in the Cockpit: A Low-Fidelity Investigation. Proceedings of the Human Factors Society 36th Annual Meeting. Santa Monica.

Tracy, K. & Moran, J.P.III (1983). Conversational relevance in multiple goal settings. In R.T. Craig and K. Tracy (Eds.), Conversational coherence: Form, structure, and strategy. Sage; Beverly Hills, CA.

Tukey, J.W. (1969). Some Graphic and Semigraphic Displays. Paper presented at the annual meetings of the American Statistical Association, August, 1969.

Tukey, J.W. (1977). Exploratory Data Analysis. Addison-Wesley; Reading.

Wainer, H. (1990). graphical Visions from William Playfair to John Tukey. Statistical Science, Vol. 5.

Walcott, C. (1972). Bird navigation. Natural History, 81.

Webster's New Universal Unabridged Dictionary (1983). Simon & Schuster, N.Y.C.

Wickens, C.D., Todd, S., & Seidler, K.S. (1989). 3-D Displays: Perception, Implementation and Applications. CSERIAC report.

Wickens, C.D. (1992). Engineering Psychology and Human Performance. HarperCollins Publishers, Inc.; New York, NY.

Wiener, E. L. (1963). On the Probability of Detection of a Signal by Multiple Monitors. Psychol. Record, 13.

Wiener, E.L. (1988). Cockpit Automation. In E.L Wiener & D.C. Nagel (Eds.) Human Factors in Aviation. San Diego: Academic Press Inc.

Wiener, E.L., Chidester, T.R., Kanki, B.G., Palmer, E.A., Curry, R.E., and Gregorich, S.E. (1991). The Impact of Cockpit automation on Crew Coordination and communication: I. Overview, LOFT Evaluations, Error Severity, and Questionnaire Data. NASA Contractor Report 177587.

Wiltschko, R., D. Nohr, and W. Wiltschko (1981). Pigeons with a deficient sun compass use the magnetic compass. Science, 214.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE May 1994		3. REPORT TYPE AND DATES COVERED Contractor Report
4. TITLE AND SUBTITLE Effects of Checklist Interface on Non-verbal Crew Communications			5. FUNDING NUMBERS NCC2-486	
6. AUTHOR(S) Leon D. Segal				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Western Aerospace Laboratories, Inc. 1611 Mays Avenue Monte Serrano, CA 95030			8. PERFORMING ORGANIZATION REPORT NUMBER A-94079	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001			10. SPONSORING/MONITORING AGENCY REPORT NUMBER NASA CR-177639	
11. SUPPLEMENTARY NOTES Point of Contact: Barbara Kanki, Ames Research Center, MS 262-3, Moffett Field, CA 94035-1000; (415) 604-0011				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified — Unlimited Subject Category 03			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The investigation described hereunder looked at the effects of the spatial layout and functionality of cockpit displays and controls on crew communication. Specifically, the study focused on the intra-cockpit crew interaction—and subsequent task performance—of airline pilots flying different configurations of a new electronic checklist, designed and tested in a high-fidelity simulator at NASA Ames Research Center. The first part of this proposal establishes the theoretical background for the assumptions underlying the research, suggesting that in the context of the interaction between a multi-operator crew and a machine, the design and configuration of the interface will affect interactions between individual operators and the machine, and subsequently, the interaction between operators. In view of the latest trends in cockpit interface design and flight-deck technology—in particular, the centralization of displays and controls—the introduction identifies certain problems associated with these modern designs, and suggests specific design issues to which the expected results could be applied. A detailed research program and methodology is outlined, and the results are described and discussed. Overall, differences in cockpit design were shown to impact the activity within the cockpit, including interactions between pilots and aircraft and the cooperative interactions between pilots.				
14. SUBJECT TERMS Non-verbal communication, Consequential communication, Interface, Cockpit design, Crew communication			15. NUMBER OF PAGES 97	
			16. PRICE CODE A05	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT	